SWEET BRIAR COLLEGE
3 2449 0405945 6

Digitized by the Internet Archive
in 2010 with funding from
Lyrasis Members and Sloan Foundation

http://www.archive.org/details/constructionofco00thom

Construction of a Cosmid Contig on Chromosome 12

Senior Honors Thesis
Highest Honors in Biology
Cassandra Thomas

This Thesis is dedicated to Dr. Davies for her numerous hours spent teaching, counseling, and encouraging me through both the good and the bad times. The effort she gave as both a teacher and an advisor went well beyond the call of duty and for this I will always be grateful. I never would have made it without her positive attitude, never-ending smile, and continuous support. Thank you Dr. Davies!! ☺

Thanks also to my parents for giving me the chance to get a wonderful education at Sweet Briar. Thanks for always believing in me and loving me.

Thanks also to Dr. Joanne Rosinski and Dr. Adele Addington for there time and effort spent helping me and supporting me.

Table of Contents

I. Abstract 1
II. Introduction 2
    What is a Contig? 2
    Birth of the Project 3
    The Human Genome Project 3
    Goals of the Human Genome Project 4
    Recombinant DNA 15
        Cosmids 15
        Yeast Artificial Chromosomes 17
    Sequence Tagged Sites 18
    Synthesis of Primers 19
    Polymerase Chain Reaction 19
    Gel Electrophoresis 23
    Labeling of Probes for Hybridization 24
    Hybridization and Non-radioactive Detection 25
    Restriction Endonucleases 27
    Restriction Enzyme Digestion 28
    Creating the Contig 30
III. Materials and Methods 31
    Recombinant DNA Preparation 31
        Growth of Eschericia coli 31
        Isolation of Cosmids from E. coli 31
    The Polymerase Chain Reation 33
        Determination of Primer Sequence 33
        Synthesis of Primers 33
        Purification of Primers 35
        Ethanol Precipitation of Primers 36
        Determination of Primer Concentration 36
        PCR Trials 37
    Oligonucleotide Hybridization 38
        Preparation of Controls for Hybridization 38
        Labeling of Oligonucleotides 38
        Control Dot Blot 38
        Hybridization 39
        Colorimetric Detection 40
        Chemiluminescent Detection 41
    Restriction Digestions and Gel Electrophoresis 42
        Restriction Enzyme Digestion of Cosmids 42
        Preparation of Agarose Gels 42
        Gel Electrophoresis 43
        Staining of DNA 43

IV. Results 45
- Synthesis of Primers 45
- Growth of E. coli and Isolation of Cosmids 46
- PCR Trials 47
- Restriction Digestions 54
- Hybridization 58

V. Discussion 63
- Synthesis of Primers 63
- Isolation of cosmids from E. coli 63
- PCR Trials 64
- Restriction Enzyme Digestions 66
- Construction of the Contig 66
- Hybridization 67

References 73

Abstract

The ultimate goal of the Human Genome Project is the individual base sequence of all 24 human chromosomes. In order to sequence the DNA, it must first be broken up into very small pieces. These small pieces, like pieces of a puzzle, are then placed in the order in which they belong on the chromosome. If these small pieces of DNA are in order then their corresponding sequences will also be in order. It is the goal of this thesis project to determine the order of approximately 200 small pieces of chromosome 12 by matching them to tags, or markers, on the chromosome.

To help in the determination of their order, pieces of DNA may be matched to tags, or markers on the chromosome. There are two ways to go about doing this. The first method is the Polymerase Chain Reaction (PCR). By using PCR, cosmid 128D6, which had been previously matched to marker D12S950, was found to span marker D12S1987. This also determined the specific location of marker D12S1987, adjacent to marker D12S928. This was unexpected as pieces of DNA this small usually only match one marker. PCR, along with restriction enzyme digestion, was also used to order a few overlapping pieces of DNA which all centered around marker D12S994.

The second method for matching pieces of DNA to markers is hybridization. All attempts at hybridization were found to be unsuccessful.

Introduction

What is a Contig?

A contig is made up of small pieces of DNA which have overlapping sequences (see figure 1). These pieces are usually recombinant DNA (DNA which has been taken from more than one organism and joined together). Depending of which kind of recombinant DNAs are used to construct the contig, the contig can have different names. The contig which was attempted in this thesis project is based on pieces of recombinant DNA called cosmids. Therefore the contig is called a cosmid contig.

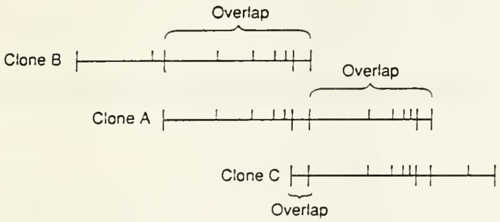


Figure 1: Assembly of a contig showing overlapping pieces of DNA(7).

Recombinant cosmids are made by breaking up a large piece of DNA (in this case chromosome 12) into smaller pieces and joining them to vectors called cosmids. The puzzle is to then put them back into the order in which they belong on the chromosome. To aid in the determination of the cosmids' order, they can be matched to markers along the DNA. Markers are much smaller than cosmids and are useful because they have easily recognizable unique sequences along the chromosome. Since both the sequence and the position of each marker is known, the location of a cosmid which matches a

marker is also known. Once the position of a cosmid is known, then the positions of any cosmids which overlap it, or share part of its sequence, are also known. This group of cosmids form a contig. Once a contig is formed the cosmids can be sequenced to determine the nucleotide sequence of the entire contig.

Birth of the Project

This thesis project originated at the Albert Einstein College of Medicine of Yeshiva University where a group of scientists led by Dr. Raju S. Kucherlapati are presently creating maps of chromosome 12 (10). All of the work which has been completed has been in collaboration with Dr. Kucherlapati's lab. Dr. Kate Mongomery, a member of Dr. Kucherlapati's lab, has been of great help in supplying cosmids and primers as well as necessary information about the chromosome.

The Human Genome Project

In 1984 a small group of molecular biologists were brought together by the U.S. Department of Energy in Alta, Utah. According to a congressional order, they were to investigate the genetic damage inherited by children whose parents had exposure to low levels of radiation and other mutagens. After talking, the group realized that the only way they could determine if genetic damage was present would be to examine an individual's DNA sequence since the majority of mutations were changes in single nucleotides. Once an individual's base sequence was obtained they would need to compare it to an intact, non-mutated genome. This interest in the individual base

sequence of humans grew rapidly into what we now know as the Human Genome Project (HGP) (17).

In 1986, James Watson proposed that the project should be centered at the National Institutes of Health instead of at the Department of Energy since genetics was traditionally the province of the NIH. The National Academy of Sciences then appointed a National Research Council Committee to study the idea. They deliberated for 14 months and then urged Congress to officially begin the HGP and to invite other nations to join in the efforts. Meanwhile, the Department of Energy, NIH, the National Science Foundation, and the Howard Hughes Medical Institute already had projects underway (17).

Goals of the Human Genome Project

The goals for the Human Genome Project for 1990-1995 were the construction of genetic-linkage maps, construction of physical maps, and the development of DNA sequencing technology. In 1993, the goals to be completed by the year 1998 included development of processes to test large numbers of individuals for genetic-linkage map markers, the development of markers which are easier to use than those available, completion of physical maps with a resolution of 100kb between markers, development of better, faster sequencing techniques, and development of the technology for high throughput sequencing systems which can integrate all steps from template preparation to data analysis (6).

Genetic-linkage maps show how genes and other recognizable pieces of DNA, called markers, are arranged on the chromosome. They are based on a phenomenon

called crossing-over which occurs when two chromosomes come close to each other (synapse) in meosis. One part of one chromosome is tightly complexed over its homologous chromosome, and the two strands break and then are rejoined (see figure 2). If crossing over happens between two genes, the alleles of the gene furthest away from the centromere are switched to the opposite chromosome. This means that genes which were originally on the same chromosome and normally inherited together will now be on separate chromosomes and will no longer be inherited together. If two genes are close together on a chromosome then it is less likely that crossing over will take place between them, and therefore more likely that they will be inherited together. However, if two genes are far apart on the chromosome then it is more likely that crossing over will occur between them and that they will not be inherited together (17).

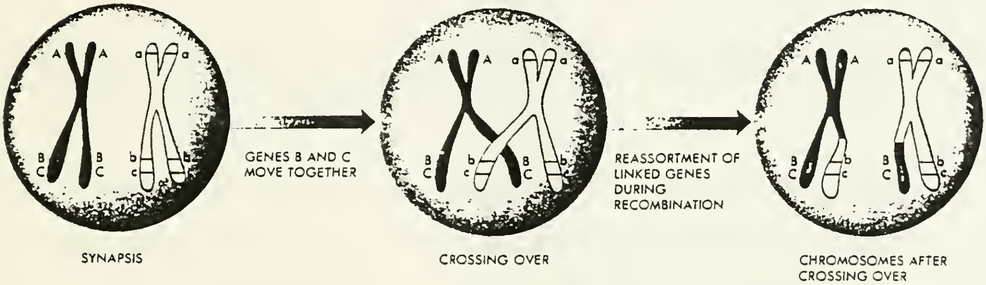

A
A
a
a
B
C
B
C
b
c
b
c
SYNAPSIS
GENES B AND C MOVE TOGETHER
CROSSING OVER
REASSORTMENT OF LINKED GENES DURING RECOMBINATION
CHROMOSOMES AFTER CROSSING OVER

Figure 2. Crossing over on a chromosome resulting in reassortment of genes (38).

Genetic-linkage maps are based on this assumption and are made by measuring the frequency with which two different traits are inherited together (see figure 3). The distances between genes and markers on a given genetic map are very long, averaging approximately ten million base pairs in length. This is approximately 5% of a

chromosome. Genetic maps are therefore not used to tell us exactly how far apart different genes are, but instead to tell us which genes are relatively close together, and which are relatively far apart (17). The unit of measurement which is used in genetic-linkage mapping is the centimorgan and is equal to 1% recombination (39). One of the goals of the HGP was to have physical maps with a resolution of 2-5 centimorgans. This goal was met and exceeded with the development of maps of the entire genome with a resolution of 0.7 centimorgans or better (23).

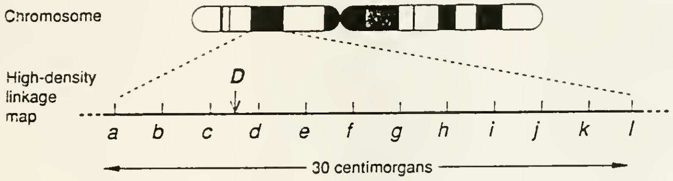


Figure 3: Example of a genetic-linkage map. This map shows the position of a gene between two markers. The position of gene D was determined through linkage analysis. Notice that the distances are measured in centimorgans rather than base pairs (7).

The information used to construct genetic maps tells us which individuals showed a characteristic and which did not. This information is usually gathered from family trees because it is both unethical and too time consuming to perform breeding experiments on humans.

Physical maps on the other hand are "derived principally from chemical analysis of the DNA molecules themselves" (17). Because physical maps rely on observing and testing the DNA directly, they are able to show not only the linear order of the genes and other segments of DNA, but are also able, at least in the more detailed maps, to show the

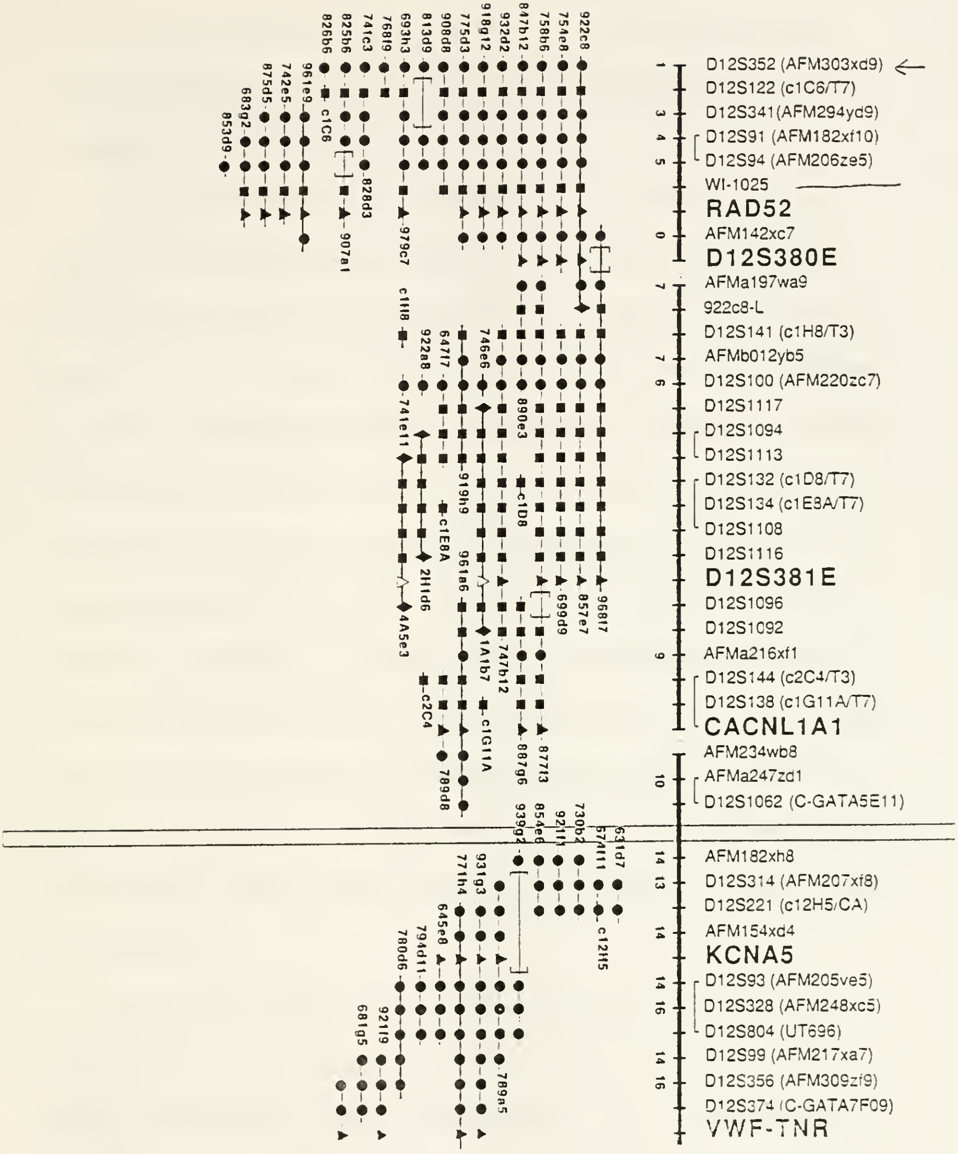


Figure 4: Physical map of the p terminus of chromosome 12. The dark vertical line represents the chromosome divided along its length by STS markers and genes (bold print). YACs are shown as shorter vertical lines overlapping one another (16).

relative distances (in base pairs) between two markers. They are formed by finding overlapping sequences in pieces of recombinant DNA and bridging these pieces all the way across the chromosome (17).

One of the first physical maps of chromosome 12 is located in The Genome Directory of the journal Nature (16) (see figure 4). This map is based on overlapping pieces of recombinant DNA called Yeast Artificial Chromosomes (YACs) which are very large and therefore span larger sections of the chromosome than most other recombinant DNAs. This is a second generation map, more detailed than the first physical map of the chromosome which was constructed by Cohen *et al.*(5) The original map had only 21 markers for the entire chromosome and the largest contig spanned only 5 of these. The second generation map includes markers and genes with an average spacing of 248 kb. The 891 YACs in the map have an average insert size of >900 kb and there are only 11 gaps along the chromosome which do not contain a contig (continuous set of overlapping DNA). The p terminus of this map is the work of Dr. Kate Montgomery of the Albert Einstein College of Medicine (AECOM). Part of the p terminus, from marker D12S352 to marker WI-1025, was used as the foundation for this thesis project to create a higher resolution map (16).

As her research continued, Dr. Montgomery generated a third generation map which extended the p terminus by seven markers and added three new markers to the D12S352 - WI-1025 section referred to above (see figure 5). Sixteen of these markers were used in this project. Figure 6 shows the relationships of the different types of maps available including genetic linkage, physical map based on a YAC and cosmids, and the

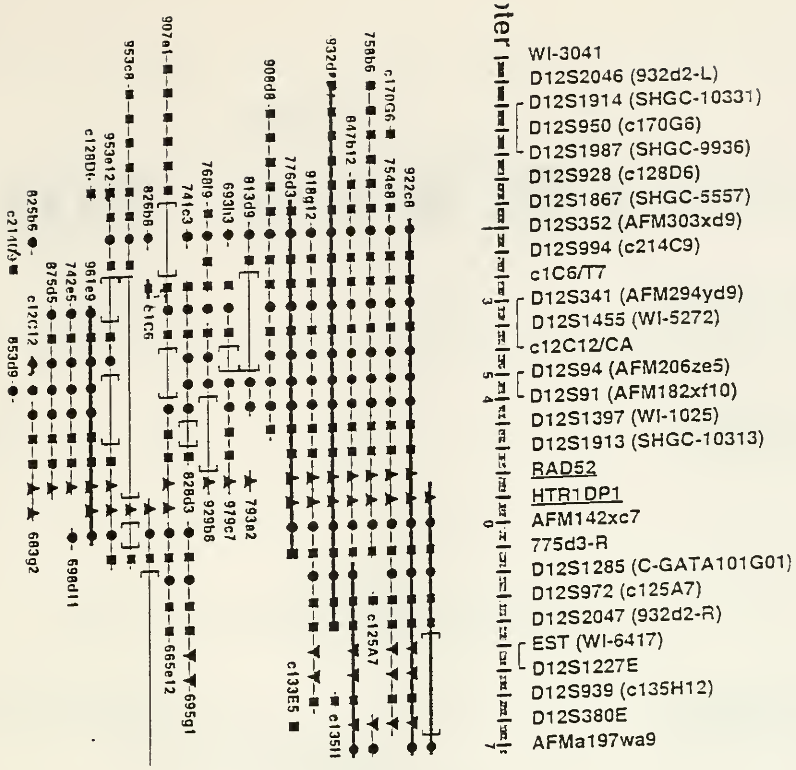


Figure 5: Physical map of the p terminus of chromosome twelve constructed by Dr. Kate Montgomery. Notice the additional markers which have been added.

individual base sequence. The process of creating maps of higher and higher resolution is critical to the eventual goal of determining the individual base sequence of the chromosomes because the low resolution maps contain pieces of DNA which are too large to sequence.

Sequencing of DNA involves using a computerized version of the Sanger technique, also called the dideoxy method (30), in which pieces of DNA are "fed" into an instrument which performs a sequencing reaction and electrophoretic analysis, and then

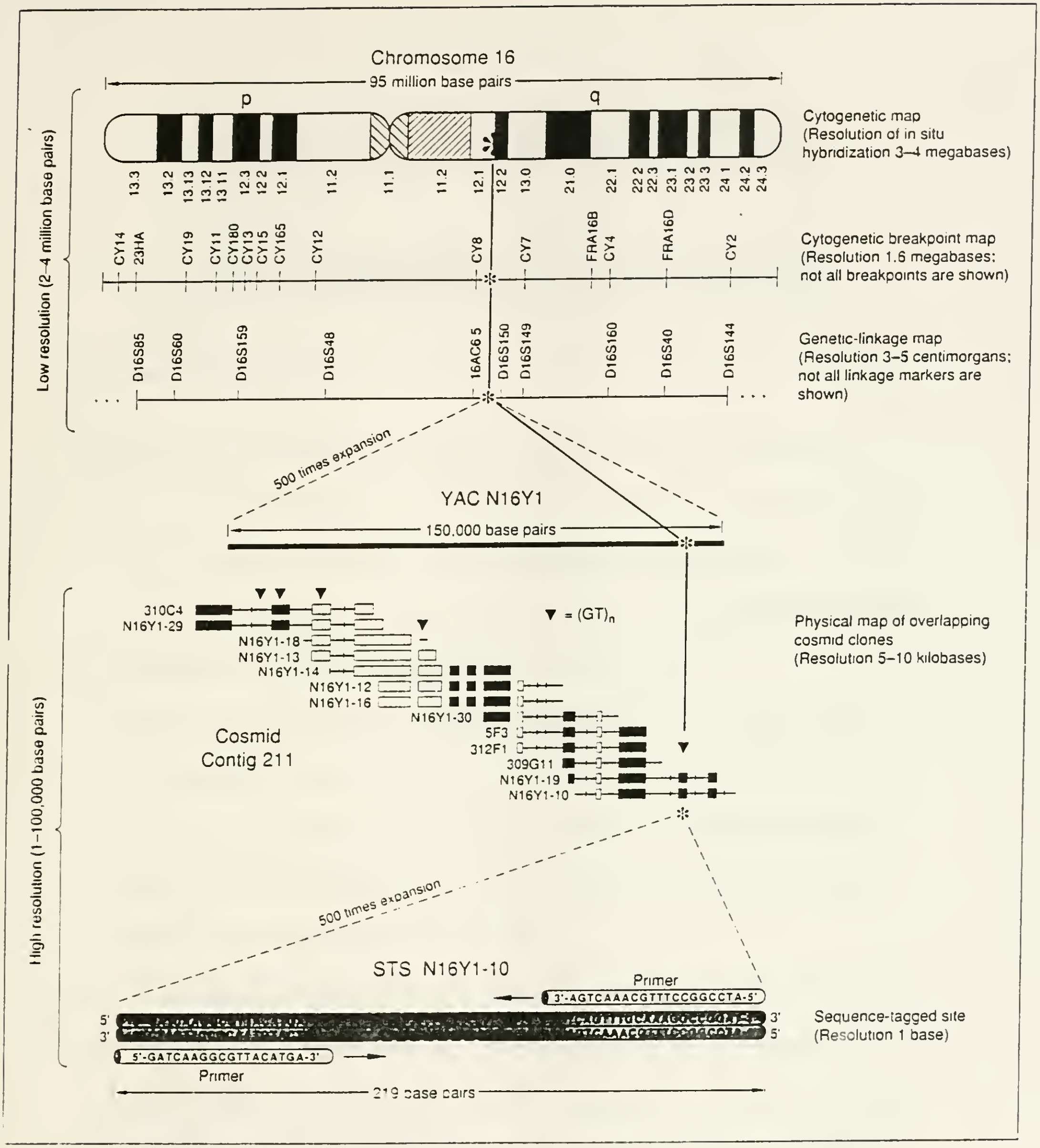


Figure 6: The relationship of the different types of maps available from the lowest resolution to the highest resolution, the actual base sequence (7).

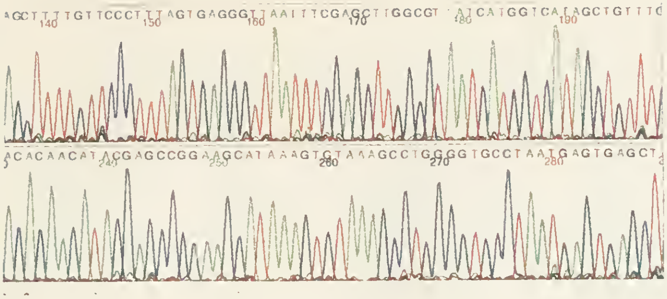


Figure 7: A section of the output of a sequencing system. Each of the four differently colored curves indicates a different dNTP (36).

gives a readout of the sequence (see figure 7). This method of sequencing uses the Klenow fragment (the polymerase part of the enzyme) of a DNA polymerase from *E. coli* to make complementary copies of the single stranded DNA to be sequenced. The DNA template, DNA polymerase, primers to initiate replication, and the four deoxynucleoside triphosphates (dNTP) are incubated along with a small amount of 2',3'-dideoxynucleoside triphosphate (ddNTP). In the simplest case, only ddNTPs of one base are added to a reaction, and four reactions, one with each of the ddNTPs, are performed. As the chain is extended by the polymerase, occasionally a ddNTP will be added to the growing chain instead of a dNTP. This causes extension of the chain to stop because of the absence of the 3` OH group.

The products of the reaction are then run on a gel in four parallel lanes (see figure 8). Each chain would be terminated at a different position because of the random addition of ddNTPs, therefore causing a band to be formed for each base of the sequence. The primers, which are labeled with a fluorescent dye cause the sequencer to detect

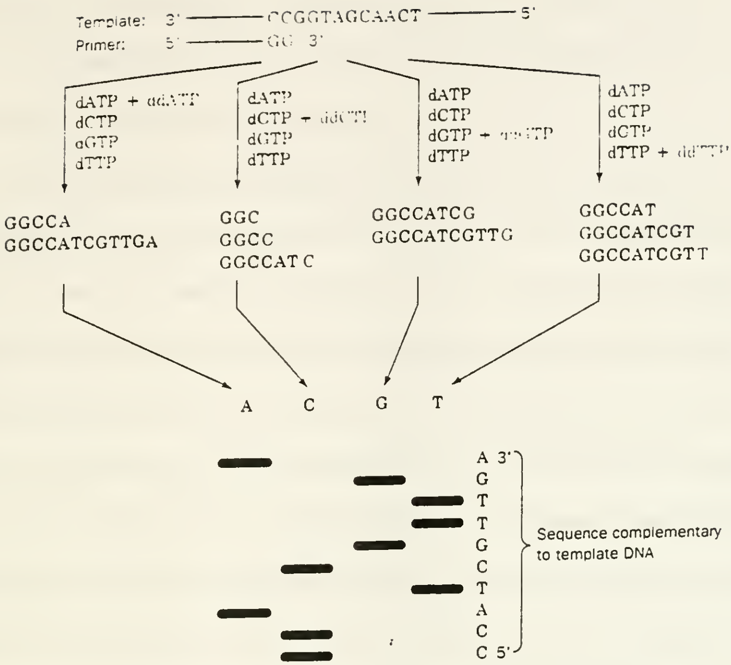


Figure 8: The addition of the four dideoxynucleotide reactions to a gel to determine the sequence of a piece of DNA (36).

the different bands as they pass through its sensors and from this a readout of the sequence is produced (36) (see figure 7).

The original map which was based on YACs could not be used for sequence determination for two reasons. First, the map has sections that are discontinuous, that is that they have no pieces of recombinant DNA matched to them. Therefore the sequence obtained would have gaps in it. Secondly, YACs are too large to analyze with a DNA

sequencer. The technology can only sequence pieces of DNA 800 bases in length or smaller (36). Even cosmids are too large to sequence on their own, but if they are digested with a restriction enzyme first and then subcloned into smaller pieces, then these smaller pieces can be sequenced directly (35).

The DNA sequencing technology of today needs major improvements if the entire three billion base pairs of the genome are to be sequenced. The first sequencing projects took several years just to sequence one gene, but by 1990 automated sequencing had dropped both in price and speed to only $5 a base and 10,000 bases a day. If these figures were to remain the same it would cost over 15 billion dollars and take one sequencer over 820 years to sequence the entire genome or it would take 820 sequencers 1 year. These numbers, however, only represent the time and money needed if there were no overlaps and if each piece of DNA was only sequenced once. In order to be accurate, the DNA must be sequenced a minimum of two times with three times being ideal. Obviously this is still too expensive and too time consuming so the technology is continuing to improve (17).

The goal of the HGP is to develop maps for and then determine the individual base sequence of each of the 24 human chromosomes. Within these chromosomes are approximately 50,000-100,000 genes as well as a great deal of inter-genic material. Together these add up to a grand total of 3 billion base pairs. It is the sequence of the genes which most people think of when we say we are going to sequence the genome. One reason for this is that knowing the sequence of the genes will give us the power to determine the causes of many genetic diseases and possibly develop a gene therapy cure for them. A second reason is that outside of the scientific community, the word gene has

almost become a generic term for DNA itself. Therefore it is only logical that many people believe that sequencing the genes themselves is the most important part of the Human Genome Project.

However, the sequence of the DNA which is found between the genes is just as important as the sequence of the actual genes. We do not understand the function of this inter-genic material, nor do we even know much about its structure. Several scientist have hypothesized that the inter-genic DNA may play a role in the control of gene expression, but there is no direct evidence of this. One very interesting characteristic of this inter-genic material is that it contains many families of repetitive sequences dispersed through out the entire genome (13). The large area that these sequences take up makes them important and interesting to study. By sequencing the genome, we will be able to determine more of this inter-genic structure as well as that of the genes. It is also very possible that the function of this inter-genic material will be determined along with that of many newly discovered genes.

Although the main focus of the HGP is the analysis of the human genome, a smaller part is centered around the development of maps and sequencing of other organism's genomes. The other genomes being sequenced or already sequenced are the bacterium *Escherichia coli* (3 million base pairs), the yeast *Saccharomyces cerevisiae* (14 million bp), the nematode *Caenorhabditis elegans* (80 million bp), *Drosophila melangaster* (165 million bp), *Mus musculus* or the house mouse (3 billion bp) (7). By determining the sequence of other genomes, scientists can practice and develop better techniques as well as use the information to help us understand how evolution took place. By studying the sequence homologies we can determine which organisms are most

closely related to one another. For instance, the human genome and that of the chimpanzee differ in only about 1% of their genes (17).

In 1989 Maynard Olsen, Leroy Hood, Charles Cantor and David Botstein proposed that all the data from physical mapping should be reported in a common "language" and that it should all be kept in a large database. They decided that each unique piece of DNA which had been placed onto a map would be identified as a sequence tagged site (STS) and would be entered into the STS database. In 1990, Johns Hopkins established the Genome Data Base (GDB) (11) which would collect, organize, and store genetic mapping information. The GDB would also be readily accessible to anyone working on the HGP and would distribute any information needed to them (17). With the advent of the World Wide Web, we are now able to access any of the data in the GDB at the touch of a button.

Recombinant DNA:

Cosmids

Cosmids are pieces of recombinant DNA which can be used to clone eukaryotic DNA in bacteria. Cosmids derive their name from the fact that they are a combination of plasmids and λ phage DNA. Cosmids contain two COS sites or cohesive ends which are derived from the λ phage DNA and are used to direct the insertion of the recombinant DNA into a viral head (see figure 9). The majority of the λ phage DNA has been removed so that there is room for approximately 35-45 kb of the DNA of interest (the human DNA) to be inserted (19).

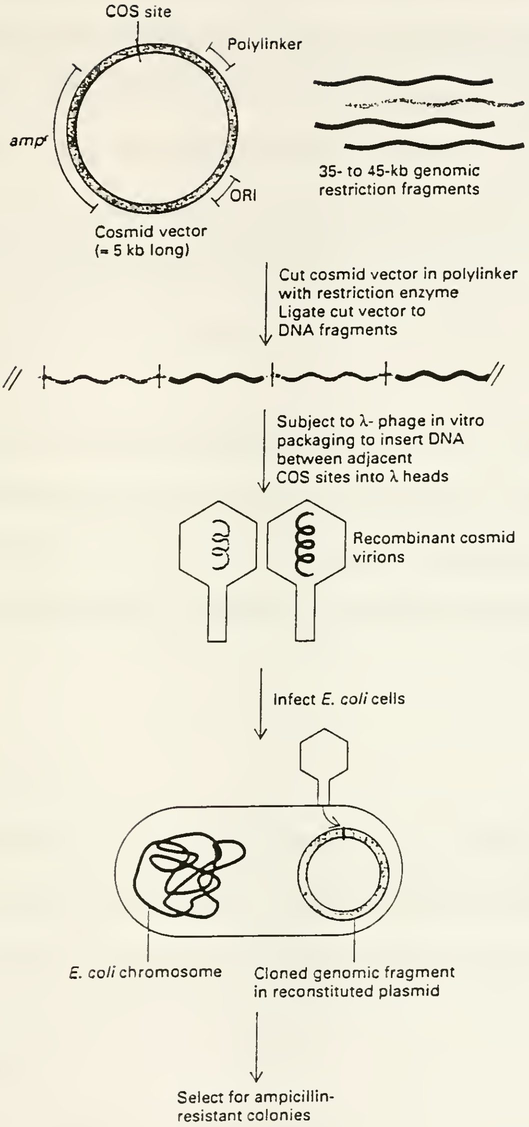


Figure 9: The construction of cosmids, their subsequent packaging into λ phage heads, and their infection of E. coli bacteria (19).

Each cosmid contains three major parts: a plasmid origin of replication which allows the cosmid to be replicated as a plasmid once inside a bacterium, restriction endonuclease sites into which the eukaryotic DNA can be inserted, as well as the eukaryotic DNA to be cloned. Most cosmids also contain antibiotic resistant genes so that clones which contain the cosmids will be resistant to an antibiotic and can be selected for. The λ phage is then used to introduce the DNA into a bacterium (39). Every new bacterial cell which arises from the transformed cell will contain at least one copy of the recombinant DNA (19).

The cosmids which were used in this project were selected by Dr. Kate Montgomery at AECOM because they were located on the end of the p-terminus of chromosome 12. She determined that they were located there by using hybridization of a mixture of primers (see hybridization section) to 2 cosmid libraries of chromosome 12 sequences (21).

Yeast Artificial Chromosomes

Yeast Artificial Chromosomes (YACs) are another type of recombinant DNA but they are very different from cosmids. Because YACs are replicated in a eukaryotic organism, they are eukaryotic chromosomes and are linear rather than circular. The construction of YACs involves incorporating the three parts of a normal yeast chromosome, the left and right telomeres and the centromere, with large pieces of DNA. The centromere is placed close to the left telomere which leaves room for hundreds of kb of human DNA to be inserted between the centromere and the right telomere (see figure 10). The large size of

YACs makes them critical to the construction of low resolution physical maps because they can span such great distances. Bob Moyzis, a key figure in the Human Genome Project, said “Without YACs we would have had little pieces of maps with no way to put them together.”(7).

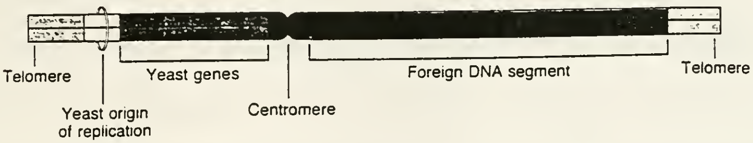
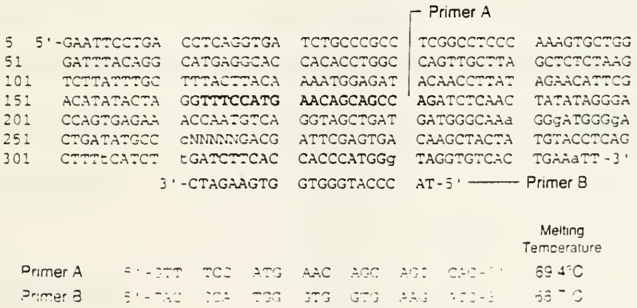


Figure 10: Structure of a Yeast Artificial Chromosome(2)

Sequence Tagged Sites

A sequence tagged site (also called a marker) is a small part of the genome, only about 200-300 bases in length, whose exact sequence is unique – it is not found anywhere else in the genome (see figure 11). STSs act as landmarks along the chromosomes to tell



```
                                                 ┌ Primer A
5    5'-GAATTCCTGA   CCTCAGGTGA   TCTGCCCGCC   TCGGCCTCCC   AAAGTGCTGG
51      GATTTACAGG   CATGAGGCAC   CACACCTGGC   CAGTTGCTTA   GCTCTCTAAG
101     TCTTATTTGC   TTTACTTACA   AAATGGAGAT   ACAACCTTAT   AGAACATTCG
151     ACATATACTA   GGTTTCCATG   AACAGCAGCC   AGATCTCAAC   TATATAGGGA
201     CCAGTGAGAA   ACCAATGTCA   GGTAGCTGAT   GATGGGCAAa   GGgATGGGgA
251     CTGATATGCC   cNNNNNGACG   ATTCGAGTGA   CAAGCTACTA   TGTACCTCAG
301     CTTTtCATCT   tGATCTTCAC   CACCCATGGg   TAGGTGTCAC   TGAAaTT-3'
                  3'-CTAGAAGTG   GTGGGTACCC   AT-5' ——— Primer B

                                                            Melting
                                                          Temperature
Primer A     5'-GTT   TCC   ATG   AAC   AGC   AGC   CAG-3'     69.4°C
Primer B     5'-TAC   CCA   TGG   GTG   GTG   AAG   ATC-3'     68.7°C
```

Figure 11: An example of an STS showing the primers from the 3' and 5' ends with the unique sequence between them(7).

us the location of a gene just as building numbers act as landmarks to tell us where a building is on a street. One of the goals of the HGP is to develop markers approximately every 100,000 bases (7).

Synthesis of Primers

Today it is possible to synthesize oligonucleotides easily and rapidly using an automated version of the phosphoramidite method of DNA synthesis. In chemical synthesis of DNA the oligonucleotide is built from its 3' end to its 5' end, unlike natural DNA synthesis which proceeds in the 5' to 3' direction. Therefore, single nucleosides of the 3' base, coupled to a solid support, are placed in the computer driven synthesizer and act as the 3' ends of the growing chains. The protecting group on the 5' end of the nucleoside is cleaved with acid and then coupled to the 3'-phosphoramidite derivative of the next nucleoside in the chain. Next, any unreacted 5' ends of the first nucleosides are capped so that they will no longer react in any of the future reactions.

These steps are then repeated for each nucleoside which is to be added to the chain. After synthesis the chain is cleaved from its solid support and purified to remove any oligonucleotides which were not successfully completed and are therefore shorter than the desired length (36).

Polymerase Chain Reaction

The polymerase chain reaction is an easy and powerful method for amplifying small pieces of the genome. PCR is carried out enzymatically in a test tube and can exponentially amplify DNA starting with only a single strand and resulting in thousands

of exact copies within a few hours. The size of the DNA to be amplified can range from 100 to 6000 bp. PCR is dependent on hydrogen bonding between a pair of primers and a piece of double stranded template DNA (22).

The primers are small pieces of DNA (approximately 20 bases in length) which are selected for their sequences. The sequences of primers must be unique within the genome and are selected from the DNA sequence located 3' and 5' to the sequence to be amplified. The sequences typically have a 50-60% G + C composition (14). The optimal annealing temperature of each primer should be as close to that of its pair member as possible. Complementarity at the 3' ends should be avoided as this increases the chances of forming primer-dimer complexes (2 primers annealed to each other) which reduces the yield of desired product. Also, sequences which contain multiples of G's and C's at either end as well as those containing palindromic sequences should be avoided as these sequences are found often in the genome and can cause mispriming to occur. If all of these conditions have been met and the primer pair selected is still unsucessful, the cause may be a secondary structure within the primer DNA itself (14).

PCR has three phases: denaturation, annealing, and extension (see figure 12). The reaction takes place in a solution containing DNA template (or target), a pair of primers, a heat stable DNA polymerase, free dioxyribonucleoside triphosphates (dNTPs) and a buffer for the correct salt concentrations. The DNA polymerase, Taq, that we use is isolated from the bacteria *Thermus aquaticus* which lives in hot springs and is therefore not denatured by the high temperatures used in this procedure (29).

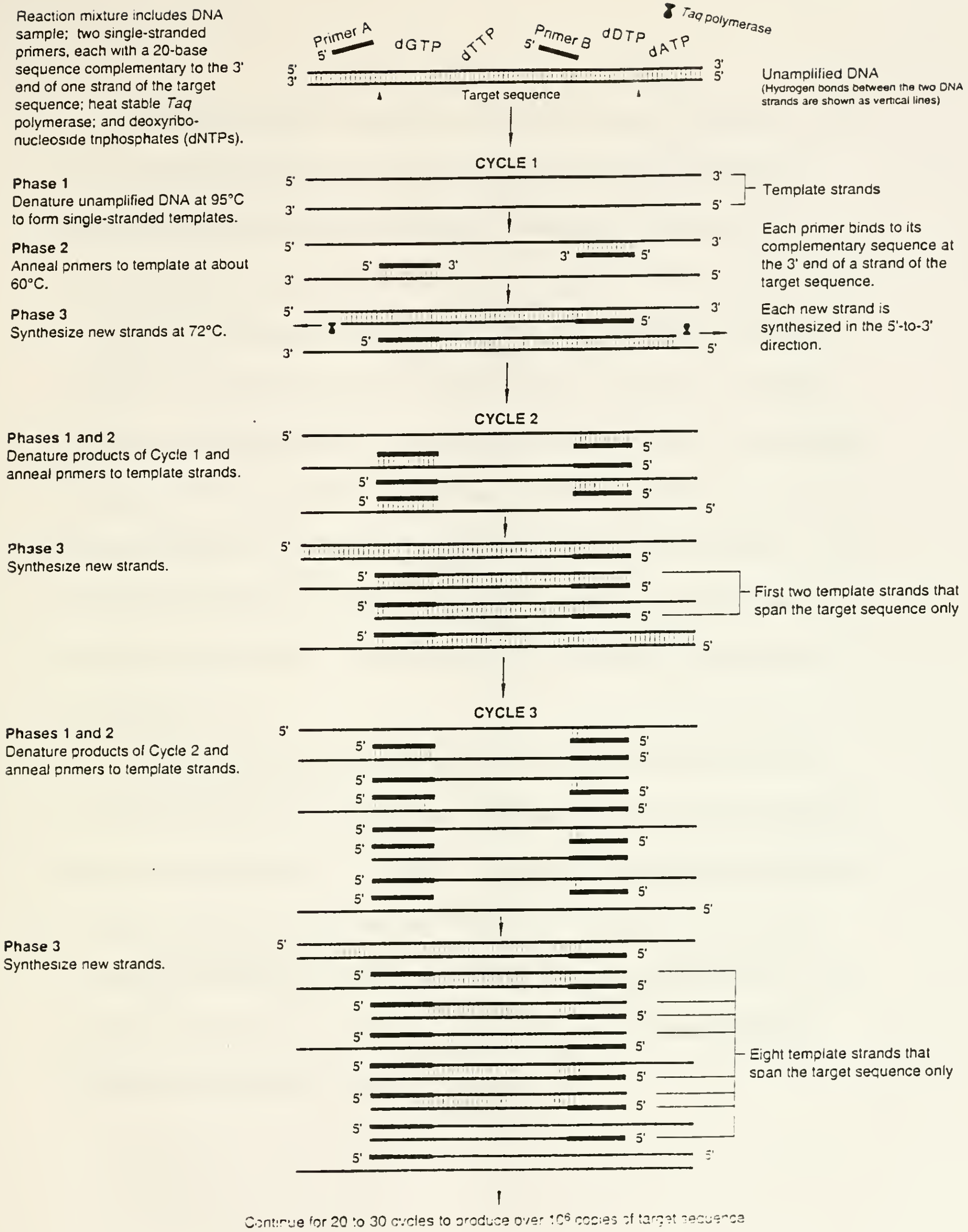

Reaction mixture includes DNA sample; two single-stranded primers, each with a 20-base sequence complementary to the 3' end of one strand of the target sequence; heat stable Taq polymerase; and deoxyribonucleoside triphosphates (dNTPs).
Primer A
dGTP
dTTP
Primer B
dDTP
dATP
Taq polymerase
5'
3'
Target sequence
Unamplified DNA
(Hydrogen bonds between the two DNA strands are shown as vertical lines)
CYCLE 1
Phase 1
Denature unamplified DNA at 95°C to form single-stranded templates.
Template strands
Phase 2
Anneal primers to template at about 60°C.
Each primer binds to its complementary sequence at the 3' end of a strand of the target sequence.
Phase 3
Synthesize new strands at 72°C.
Each new strand is synthesized in the 5'-to-3' direction.
CYCLE 2
Phases 1 and 2
Denature products of Cycle 1 and anneal primers to template strands.
Phase 3
Synthesize new strands.
First two template strands that span the target sequence only
CYCLE 3
Phases 1 and 2
Denature products of Cycle 2 and anneal primers to template strands.
Phase 3
Synthesize new strands.
Eight template strands that span the target sequence only
Continue for 20 to 30 cycles to produce over 10^6 copies of target sequence

Figure 12: The Polymerase Chain Reaction(7)

During the first phase, denaturation, the reaction mixture is heated to 94°. The heat breaks open the H-bonds of the target DNA and produces two single stranded templates. Then during phase two, annealing, the solution is cooled to the correct annealing temperature and the primers anneal or hydrogen bond to their complementary sequence on the templates. During phase three, extension, the solution is heated to 72°, the optimal temperature for the Taq DNA polymerase and the primers are extended to form two new double stranded DNA complexes. After this cycle has been accomplished it is repeated. The two double stranded complexes are denatured, primers are annealed to them, and extension is repeated with the result being four double stranded helixes. These phases are repeated approximately 30 times so that there is an exponential increase in the number of template strands. The amplified DNA can then be detected through the use of gel electrophoresis.

The uniqueness of the sequence being amplified and its primers allows the amplification of the unique marker region even if DNA from the entire human genome is placed into the reaction solution. Therefore, if a template DNA is placed into solution and a product is amplified, then the template DNA must have contained the sequence of interest. It is this idea which allows the use of PCR for developing physical maps. The PCR that I performed used a cosmid from chromosome 12 as the template DNA and primers for 16 STSs from the p arm of chromosome 12.

Gel Electrophoresis

Gel electrophoresis is a technique in which DNA is separated and then identified based on its size. The DNA is placed into an agarose gel in a buffered solution and a current is run through the gel. The negatively charged DNA then migrates towards the positive electrode due to the electrostatic attraction. The DNA, however, gets trapped in the matrix of agarose. Small pieces are able to move through the gel faster than the large pieces which are more easily retarded by the gel. One can determine the size of the DNA of interest by comparing the distance it has traveled with respect to a DNA of known size run concurrently. This is because the rate that a fragment travels is approximately "inversely proportional to the logarithm of its length" (7).

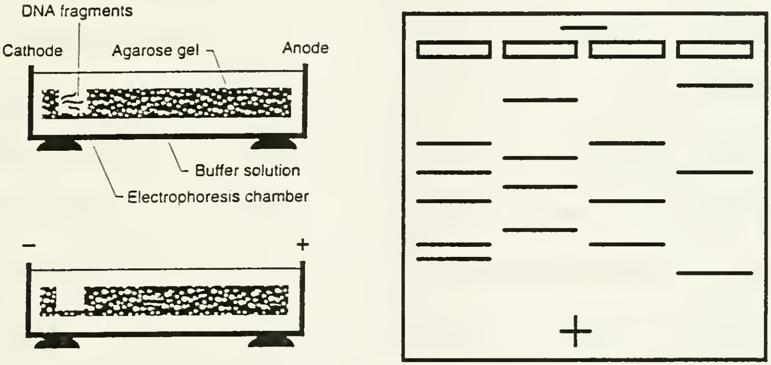


Figure 13: (a) Side view of an agarose gel before and during electrophoresis showing the migration of DNA towards the anode. (b) Top view of a similar gel(7).

The presence of double stranded DNA within the agarose gel is detected by the addition of ethidium bromide either directly to the gel, or by soaking the gel in a dilute

solution after electrophoresis. The ethidium bromide intercalates between the bases of the double helix and fluoresces visible light when exposed to ultraviolet light.

Labeling of Probes for Hybridization

A probe is a small piece of DNA which has a complementary sequence to the target DNA and has a "label" or some method of detecting its presence. In this project, the probes used were PCR primers which were labeled with psoralen biotin (see figure 14). Psoralens are photoreactive molecules which covalently attach to nucleic acids in the presence of near UV light. They undergo a two step reaction in which they intercalate into DNA in the absense of light and form covalent bonds with pyrimidines at their 3,4 and 4',5' double bonds when exposed to 320-400 nm light (28).

In this case, biotin and a spacer arm are attached to the psoralen (15,18) so that the monoadduct (DNA bound at only one end of the psoralen molecule) produces a product in which the psoralen biotin complex sticks out from the DNA and may be detected by the streptavadin-alkaline phosphatase enzyme conjugate (see figure 15). The diadduct product however produces a crosslink between two complementary strands and prevents the helix from being denatured. This diadduct product is still capable of being detected by the streptavadin-alkaline phosphate, but is not useful since the probe must be single stranded in order to bind the target (28,18). The primers that were used as probes in these experiments were denatured and then placed in solution with psoralen-biotin and exposed to UV light. Denaturation before the addition of UV light should generate the monoadduct product rather than the diadduct.

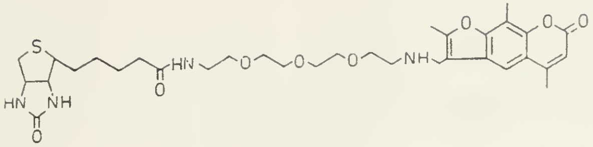


Figure 14: The structure of the psoralen-biotin label. The 3 rings to the right are the psoralen molecule with the long chain spacer arm and the 2 ring biotin attached at the left(18).

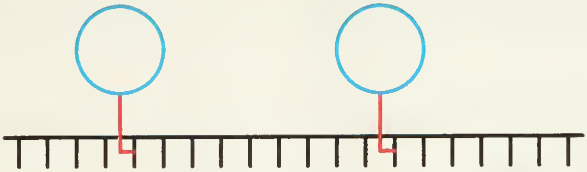


Figure 15: A typical oligo probe labeled with psoralen biotin. The red lines represent the psoralen and spacer arm attached to a pyrimidine with a blue biotin molecule "sticking out" away from the backbone of the DNA.

Hybridization and Non-radioactive Detection

Hybridization is a second method for detecting whether or not a cosmid contains an STS. During hybridization, the target DNA sequence, in this case a cosmid, is denatured to its single stranded state and then placed on a nylon membrane (9) and attached covalently by either heating the membrane to 80°C or by exposure to UV light (31). The membrane is then placed into solution with a labeled probe and incubated at the correct annealing temperature. The correct annealing temperature can be found by equations 1 or 2 (1,3). When a labeled probe comes in contact with a complementary sequence of the target DNA bound to the membrane it forms H-bonds and renatures to its

$$T_M = 81.5°C + 16.6 (\mathrm{Log10} [\mathrm{Na+}]) + 0.41 (\mathrm{GC}\%) - (600/\#\mathrm{bp\ in\ duplex})$$

Equation 1: This equation is used for the determination of TM for DNA duplexed with DNA. Correct annealing temperature is 20°C less than the T_M. TM is defined as the "temperature at which 50% of the oligonucleotide and its perfect complement are in duplex." (1)

$$Td = 2 (\#AT\ bp) + (\#GC\ bp)$$

Equation 2: This equation is used for the determination of TM for oligonucleotide probes. Correct annealing temperature is 20°C less than the Td. Td is defined as the "temperature at a particular salt concentration (0.9 M = 5X SSC) at which 50% of short oligos (14-20 mer) and their perfect filter-bound complements are in duplex.(31)

double stranded state. The psoralen-biotin then acts as a flag, sticking up from the double stranded DNA to mark the location of the target-probe complex. An analog to this process is shown in figure 16.

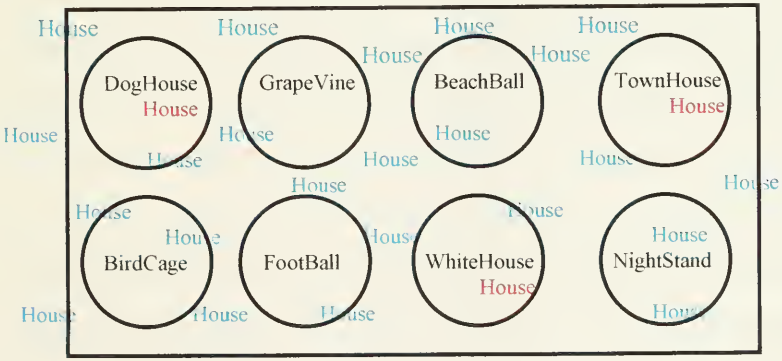


Figure 16: A Dot blot of cosmids in solution with a probe. The words inside of the circles represent the sequences of the cosmids. "House" is the probe in solution and "House" is a probe which has found its complementary sequence and hybridized.

Streptavadin-alkaline phosphate is then bound to the biotin of the target-probe complex and is reacted with a substrate to give either a colored product which can be observed on the membrane or a luminescent product, whose location can be detected by exposure to x-ray film for 3 or more hours.

Restriction Endonucleases

Restriction endonucleases, commonly called restriction enzymes, are molecular scissors which can recognize a specific base sequence of DNA and cut the double helix. Restriction enzymes occur naturally in bacteria as part of a restriction-modification system which destroys any foreign DNA which enters the host cell thereby protecting the host (19).

There are three types of restriction endonucleases. Type I restriction enzymes recognize a sequence and then cut the DNA at what is thought to be a random site located at least 1000 bp from the recognition sequence. Type III restriction enzymes perform a similar function, but cut approximately 24 to 26 bp from the recognition sequence. Neither types I or III are very useful in recombinant DNA technology. Type II restriction enzymes can be used in recombinant DNA technology because they recognize specific sequences within the DNA and cut the double helix at the recognition or restriction site. Restriction sites are usually four to eight bases long and are often palindromic (read the same forward or backwards).

The most helpful of restriction enzymes are those which make a staggered cut within the DNA, producing "sticky ends." (see figure 17) These sticky ends are single stranded regions which allow any two DNAs which have been cut with the same

restriction enzyme to bind together transiently. This means that, if they have been cut with the same restriction enzymes, any two DNAs from any two species can be combined to form recombinant DNA (19).

Restriction Enzyme Digestion

Restriction enzymes can be used to create recombinant DNA such as plasmids and cosmids, but they can also be used simply to cut DNA. As stated earlier, the physical maps or contigs are based on overlapping fragments of DNA. These fragments are created from digesting, or breaking apart, larger pieces of DNA.

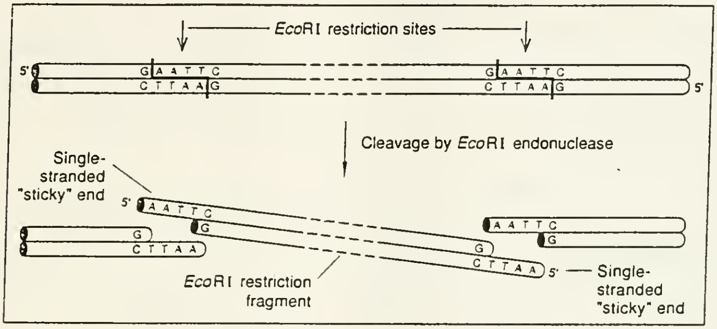
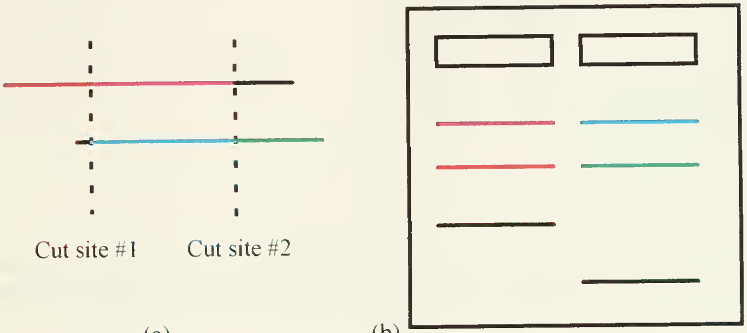


Figure 17: Cleavage of a DNA strand by *Eco*RI showing both the palindromic sequence (GAATTC) and the "sticky" ends created at each cut site (7).

Once it has been determined that two cosmids or other pieces of recombinant DNA contain the same marker, the next step is to determine how much of their sequences they share in common with one another. This is done by digesting both cosmids with the same restriction enzyme. The resulting fragments are run through gel electrophoresis to separate the fragments according to size. When the cosmids are run next to each other,

fragments of the same size will travel the same distance and can be compared by direct observation of the gel. The simple fact that two fragments are the same size does not mean, however, that they contain the same sequence. It could be that they contain completely different sequences but happened to have the same distance between to restriction sites (see figure 18).

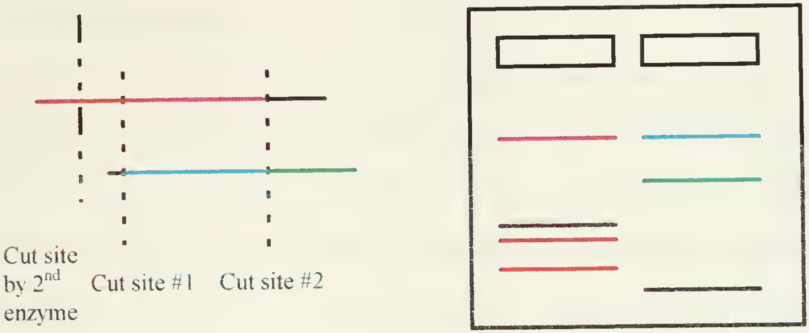


(a) (b)

Figure 18: (a) Two pieces of DNA which share the same sequence (blue and purple bands) and therefore share restriction sites resulting two bands which travel the same distance in a gel (b). The red and green bands also travel the same distance, but are not complementary sequences.

In order to help us determine if two fragments of the same size contain the same sequence we can repeat the digestion, this time using two restriction enzymes (34). This procedure is called a double digestion (see figure 19). If one of the two bands is cut by this second restriction enzyme while the other band remains intact, then we can say that the two bands did not contain the same sequence. If, however, the two bands are either both cut by the second enzyme or both remain uncut by the second enzyme then we know that the bands are more likely to contain the same sequence. Notice that this

does not give a positive answer to our question. We could continue to add restriction enzymes to remove any uncertainty, but in practice the use of more than two enzymes seems generally unnecessary (8,33).



(a) (b)

Figure 19: Double digestion. The red and green bands can now be differentiated. The red band contains the cut site (a) and is therefore digested into two smaller bands (b) while the green band remains intact.

Creating the contig

Once it has been determined that two cosmids contain the same marker and that they share restriction bands with one another, it is time to try and piece together a map. This is done by ordering the common fragments within each cosmid so that the end of one cosmid overlaps the end of the other cosmid (see figure 1). Fragments which did not overlap or fragments within an overlapping section can not always be ordered until more information is obtained from other cosmids which also overlap the area.

Materials and Methods

Recombinant DNA Preparation:

Growth of Escherichia coli

Kanamycin resistant *Escherichia coli* containing cosmids from the p-arm of chromosome 12 were obtained from Dr. Kate Montgomery at the Albert Einstein College of Medicine. Cultures were plated and grown on 2xYT agar (20) (16 g/l bacto-tryptone, 10 g/l bacto-yeast extract, 5 g/l NaCl pH 7.0, 15 g/l bacto-agar) supplemented with 20 µg/ml kanamycin at 37°C for a minimum of 12 hours. Plating was repeated until single colonies were established. Culture flasks containing 30 ml of sterile 2xYT medium and kanamycin at 20 µg/ml were inoculated with a single colony of *E. coli* and incubated at 37°C on a shaking platform overnight. The shaker was set at 300 RPM, which was sufficient to clear the bottom of each flask and thus ensure adequate aeration.

Isolation of Cosmids from Escherichia coli

Long term storage of an 800 µl aliquot of each of the liquid cultures was performed by freezing in 20% glycerol at -80°C (20). The procedure that was used for isolation of the cosmid DNA is a modification of the alkaline lysis method (20). The remainder of the liquid culture was centrifuged in sterile 50 ml tubes at 850 xg for 20 minutes. If supernatant was cloudy, centrifugation was repeated until supernatant was crystal clear. The pellet was then re-suspended by vortexing in 1.5 ml prep solution 1 (50 mM glucose, 25 mM Tris (pH 8.0), and 10 mM EDTA). Two volumes prep solution 2

(0.2 M NaOH and 1% SDS) were added and the solution was inverted gently to mix. One and one half volumes prep solution 3 (300 ml 5 M KAc, 57.5 ml glacial acetic acid, and 145.5 ml water, final pH 4.8) were added and the solution was inverted gently to mix.

The solution was then placed in the Sorvall RC-5B refrigerated high speed centrifuge and centrifuged for 15 minutes at 17,000 xg. The supernatant was filtered through Kimwipes into clean 50 ml centrifuge tubes. One half volume of isopropanol was added to the tubes which were then covered with parafilm, inverted gently to mix and allowed to precipitate at room temperature for 30 minutes. The solution was centrifuged in the Sorvall for 30 minutes at 17,000 xg. The supernatant was poured off carefully and the pellet air dried.

The pellet was then re-suspended in 400 µl sterile TE (10 mM Tris pH 8.0, 20 mM EDTA) and transferred to 1.5 ml microcentrifuge tubes. Two µl of 10 µg/µl RNase A and 2 µl of 1000 units/ml RNase T1 were added to the cosmid solution which was then incubated at 37°C for one hour. The solution was then brought to 10 mM EDTA and 0.2% SDS by the addition of 8.1 µl of 10 mM EDTA and 8.1 µl of 0.2% SDS. The solution was then transferred to a commercially prepared microcentrifuge tube containing Phase-Lock Gel which was obtained from 5 Prime → 3 Prime, Inc. (12). An equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added to each tube. The mixture was inverted several times to mix and was then centrifuged at 1400 xg for 2 minutes. The cosmid solution was then removed, brought to 0.3 M NaOAc by the addition of 150 µl of 1 M NaOAc, and precipitated with 2.5 volumes of ethanol. The solution was frozen at -80° for a minimum of two hours.

After freezing, cosmids were centrifuged at 1400 xg for 10 minutes. The supernatant was discarded and the pellet was washed with 1 ml 70% ethanol/30% TE. Pellets were allowed to air dry and were then re-suspended in 90 µl TE by gentle mixing. Cosmids were stored in solution at -20° C.

The Polymerase Chain Reaction:

Determination of Primer Sequence

Initial aliquots of 16 pairs of primers from a small section of chromosome 12 near the P-arm were obtained from Dr. Kate Montgomery of the Albert Einstein College of Medicine for use in the construction of the cosmid contig. When initial aliquots were expended these same 16 pairs were synthesized on a Pharmacia Biotech Gene Assembler® automated DNA synthesizer. The sequences of these primers were obtained from the Genome Data Base (16) which was accessed on the World Wide Web using the web browser Netscape Navigator (see table 1).

Synthesis of Primers

The Gene Assembler® from Pharmacia Biotech is operated and controlled through a computer software program. Specifications for each synthesis and the sequence to be synthesized are entered into the computer by the user. The program then directs the machine's pumps and valves to deliver the correct amount of amidites and other reagents to a support cassette at the correct time in order for DNA synthesis to occur. Once

synthesis is started, the machine is fully automated and requires no human intervention (26).

Table 1: Primers and their nucleotide base sequence.

Primer	Sequence
D12S2046F	GGATGGATAGCAGGGTCTCA
D12S2046R	GGCCGTAAGTGCAGGAGAT
D12S1914F	ATCCTCTGATGAAGCCATGG
D12S1914R	ACCCCAATGTCTGCTGTTTC
D12S950F	ACCAGGATTGATGAGGCTAC
D12S950R	ACTAGACACTGGTCCTCAG
D12S1987F	CTGAGATCATTGTGGAAAGAAGC
D12S1987R	CTACCCCAAAACTTACAAAAATGA
D12S928F	AGCTTGAGAGCTTGTCTGCG
D12S928R	AGCTGCTTATGAGAGTGTTG
D12S1867F	TTCTCCTTCATAGTGACTGTCC
D12S1867R	CCAATAAGAGCTTGGTTAGATG
D12S352F	TGGACCAGGGTGATGG
D12S352R	AGCAGGAGTCTGACGTTTTG
D12S994F	GTGTGTATGGCCTTTTGGTG
D12S994R	CCCTGAACATCCATGAAGAC
c1C6F	CCTCCGTCTCCAAATTACTAGG
c1C6R	TCAGTTAAACCGCAGGAAGG
D12S341F	TATCCAAGCCCACCCT
D12S341R	ATCTTTTACTGTTATGATGAACACA
D12S1455F	ATTTGTTAGGCAGAACCAGAGC
D12S1455R	AGTATGGAGGCCCCAGTAGA
c12C12F	GTAAATTGCTACAGGGTCATTGT
c12C12R	AATGAGAGGAGTAGTTCAGTGTCA
D12S94F	AGGTGGGAGGTTCCCTTA
D12S94R	TTGCCTAGCATCGTTGAAAA
D12S91F	TTCACAACAGCCAATGGTAG
D12S91R	TTCTCAAGGTTCGTCCATGT
D12S1397F	AAGGGGAAAATCTGTCTGCT
D12S1397R	AATTATTTCCTGAAACATTTTATGC
D12S1913F	GCTTTTTTTGGAACTTTTTAAACG
D12S1913R	ATATTAGCAGCTTTGACTGCAGC

All amidites, reagents, and support cassettes were obtained from Pharmacia Biotech and kept under pressurized argon when on the machine. The machine's pump was calibrated and amidites, tetrazole, and all solvents were purged to fill all lines with fresh solutions prior to each synthesis. For each synthesis, a solid support cassette with one nucleotide base which matched the 3' end of the primer to be synthesized was placed into the machine. All syntheses were run with a final detritylation. The trityl group or DMTr, is used to block the 5' end of each nucleotide so that it is not reacted before the proper time. By performing a final detritylation, the last DMTr group is removed and allows the primer to react in the most unencumbered manner possible (26).

Purification of Primers

When synthesis was complete, the cassette was removed from the machine and centrifuged briefly to remove the excess acetonitrile. The cassette was then placed in a clean microcentrifuge tube with of 1 ml of 30% ammonia solution and centrifuged to saturate the cassette with the ammonia solution. The cassette was placed into a test tube in a sealed jar containing a small quantity of 30% ammonia solution and incubated at room temperature overnight. The cassette was removed, placed into a new microcentrifuge tube, and centrifuged to collect the free primers. Dissolved primer/ammonia solution from both microcentrifuge tubes was then combined and brought to a final volume of 1 ml with sterile water if necessary. Primers were then purified using a NAP molecular sieve column (26) purchased from Pharmacia which removes ammonia, salts, released protecting groups and oligos less than ten bases in length. The column was washed with sterile distilled water several times. The

primer/ammonia solution was then added to the NAP column. The primers were removed by the addition of 1.5 ml of sterile water to the column and were collected in a clean microcentrifuge tube.

Ethanol Precipitation of Primers

The primer solution was then split into four sterile microcentrifuge tubes (26). To each tube, 1 M NaOAc was added to 90 mM; 1 M $MgCl_2$ to 9 mM, and then 2.1 volumes of ethanol were added. The solution was frozen at -80°C for a minimum of two hours. The tubes were then centrifuged for 10 minutes at 1400 xg. The supernatant was removed and pellets were washed in 1 ml 70% ethanol/30% TE and re-centrifuged to remove any salts. The supernatant was removed and the pellet was air dried. Pellets were then dissolved in 400 µl sterile TE.

Determination of Primer Concentration

The concentrations of the primers were then determined by measurement of their OD_{260} in a Beckman DU-640 spectrophotometer. A 1:250 dilution of primer solution was prepared and placed into a quartz cuvet. DNA concentrations were then determined using the equation: $1.000\ OD_{260} = 33\ \mu g/ml$ (20). The result was then multiplied by the dilution factor of 250.

PCR Trials

PCR was performed using known primers to amplify unique markers on Chromosome 12 (16). Reactions were performed in 20 μl total volume and were run in the Ericomp PowerBlock I thermocycler. Each reaction was performed according to the package insert supplied with the DNA polymerase. Each reaction contained 1 mM Tris-HCl, pH 8.3, 5 mM KCl, 200 μM of each dNTP, 1.5-2.5 mM $MgCl_2$ (dependent on the optimal annealing conditions for each primer pair), 100-200 ng of DNA template, and 0.5 units *Thermus aquaticus* DNA polymerase (*Taq*). Reaction buffer (10x concentration: 10 mM Tris-HCl, 50 mM KCl) and *Taq* enzyme were purchased from Perkin Elmer. Primer concentrations varied, but averaged 20 ng/reaction. Sterile double-distilled water was used to bring the final volume to 20 μl.

Conditions for PCR were: initial denaturation at 94°C for 2 minutes, then 30-35 cycles of PCR, each consisting of denaturation at 94° for 30 seconds, annealing at the primer's optimal temperature (see table 4) for 30 seconds, and extension at 72°C for 30 seconds. These were followed by a final extension at 72°C for 5 minutes (11).

In all reactions, controls were run with one or more of the following as template DNA: total human DNA (K562) purchased from Gibco BRL, chromosome 12 DNA from the human/mouse hybrid cell line GM10868, YAC 932d2, or a previously mapped cosmid. Following this, single cosmids as well as pools of cosmids were run with different primers. PCR products were observed using gel electrophoresis in 1.5% agarose gels supplemented with 0.3 μg/ml ethidium bromide.

Oligonucleotide Hybridization:

Preparation of Controls for Hybridization

Control target DNA (pUC 19), control labeled oligo (a 20mer complementary to pUC 19), psoralen biotin, Rad-Free blocking powder, and Nytran membrane were all obtained from the Schleicher and Schuell RAD-FREE oligo labeling and hybridization kit. All procedures for hybridization and its preparation were obtained from Schleicher and Schuell. Control target DNA and the control oligo probe were reconstituted from the lyophilized state with sterile dH_2O.

Labeling of Oligonucleotides

Ten µl of 100 ng/µl of primers were boiled for ten minutes to denature any self-annealing and then placed directly on ice to maintain their single stranded state. Ten µl of primers were then placed in a clean 96-well plate which was placed on top of an ice water bath in a darkened room. Psoralen biotin reagent (1 µl) was added to the primers and the plate was irradiated for 1 hour under the UV lamp provided with the Rad-Free oligo labeling and hybridization kit to crosslink the psoralen to the primers. Labeled primers were taken up in 90 µl TE and then placed in clean microcentrifuge tubes and extracted twice with 1 ml H_2O-saturated n-butanol to remove any unlinked psoralen.

Control Dot Blot

Seven serial dilutions of the control pUC19 DNA were made in 5X SSC (750mM NaCl, 75 mM TrisNaCitrate$\cdot$$2H_2O$). Each tube was then placed into a boiling water bath

for ten minutes and then placed immediately into an ice water bath for five minutes. Five microliters of each tube were then placed onto a Nytran membrane which had been pre-wet in 5X SSC. The membrane had been divided into 1 cm x 1 cm squares and the DNA was pipetted onto the membrane in the following quantities: 30 ng, 6 ng, 1.2 ng, 240 pg, 48 pg, 9.6 pg, and 1.9 pg of DNA. Duplicate applications of the DNA were performed. The membrane was then placed into the UV Stratalinker and exposed to 12,000 μj of radiation to crosslink the DNA to the membrane. The membrane was stored at -20°C.

Hybridization

The dot blot membrane was wet in 1X TBS (50 mM Tris, 150 mM NaCl, pH 7.5) and then washed in 5X SSC prior to hybridization. The blot was then placed into a hybridization bag with 10 ml pre-hybridization solution (5X SSC, 5% Rad-Free Blocking Powder, and 1.0% SDS) for every 100 cm^2 of membrane and the bag was sealed. The bag was placed in a water bath and incubated for 1 hour at 42°C. Final concentration of the probe in hybridization solution was 25 ng/ml. Concentrated probe solution (750 ng/ml) was placed in a boiling water bath to denature for 10 minutes followed immediately by immersion in an ice water bath for five minutes. The concentrated probe solution was then placed in hybridization buffer (10 ml for every 100 cm2 of membrane). The bag was opened and the pre-hybridization buffer discarded. Fresh hybridization solution with probe was added to the bag and it was resealed. The blot was then incubated in a water bath at 42°C for 1 hour with gentle agitation approximately every 15 minutes.

The blot was then removed from the bag and washed in 500 ml wash solution I (2X SSC, 0.1% SDS) for five minutes at room temperature with gentle rocking. This wash was repeated once. The blot was then transferred to wash solution II (0.1X SSC, 0.1% SDS) which had been pre-heated to 42°C and was washed for five minutes. This wash was also repeated once.

The blot was then placed in a new hybridization bag with 30 ml of detection blocking solution (1X TBS, 3% Rad-Free Blocking Powder, 0.1% SDS) for every 100 cm^2 of membrane and the bag was sealed. This was placed on a shaker platform and incubated at room temperature for 3 hours with gentle agitation. The blocking solution was then removed and fresh blocking solution containing streptavadin-alkaline phosphatase at a dilution of 1:1650 was added. The blot was incubated at room temperature for 1 hour with gentle agitation. The blot was then washed in wash solution III (1X TBS, 0.1% SDS) for 10 minutes at room temperature with gentle rocking. This wash was repeated twice. The blot was then washed in 1X TBS for 10 minutes at room temperature with gentle rocking.

Colorimetric Detection

A Rad-Free BCIP/NBT tablet, purchased from Schleicher and Schuell, was placed in 30 ml dH_2O in a sterile 50 ml centrifuge tube and shaken vigorously until dissolved. This was then placed into a hybridization bag with the blot after the final washing and was sealed. The blot was placed on a shaking platform and covered with aluminum foil. This was incubated at room temperature overnight with gentle agitation.

Chemiluminescent Detection

During hybridization, a piece of Schleicher and Schuell, Rad-Free Lumi-Phos®530 chemiluminescent substrate sheet within its sleeve was placed into a side seal reaction bag. This assembly was placed into an imaging cassette and incubated at 37°C while the dot blot was being processed. After the dot blot membrane's final wash, the substrate sheet sleeve was removed from the imaging cassette and the side seal bag and was laid open on a piece of clean absorbent paper. The blot was placed nucleic acid side up on the substrate sheet and the plastic sleeve placed back on top of the blot. A glass pipette was rolled across this assembly in order to ensure good contact between the blot and substrate. The substrate sheet sleeve, containing the dot blot and substrate sheet, was then placed back into the side seal reaction bag and the sides were sealed. This entire assembly was then placed back into the imaging cassette along with a piece of unexposed x-ray film which was laid directly on top of the side seal reaction bag. The cassette was then incubated at 37°C for 3 hours to overnight.

Restriction Digestions and Gel Electrophoresis:

Restriction Enzyme Digestion of Cosmids

Five restriction enzymes; *Eco*R I, *Bam*H I, *Hin*d III, *Nco* I, and *Apa*L I; were originally used to digest cosmids. These enzymes were supplied by New England Biolabs and were used according to the manufacturer's instructions. All reactions were carried out in 20 µl volumes. Optimal conditions for *Eco*R I reactions were: NEBuffer *Eco*R I (50 mM NaCl, 100 mM Tris-HCl, 10 mM $MgCl_2$, 0.025% Triton X-100, pH 7.5); 5 µg cosmid DNA, and 1 U *Eco*R I. Sterile water was used to bring the solution up to the correct volume.

Optimal conditions for *Bam*H I reactions were: NEBuffer *Bam*H I (150 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9), 5 µg cosmid DNA (~ 1 µg/µl), and 1 U *Bam*H I. Sterile water was added to bring solutions to correct volumes. If both *Eco*R I and *Bam*H I were used then the conditions were as follows: 2 µl NEBuffer *Eco*R I, 5 µg cosmid DNA, and 1 U of each enzyme. Sterile water was used to bring the solution to correct volume. Reaction conditions for all solutions were incubation at 37° for 1-3 hours. The DNA was then examined using gel electrophoresis.

Preparation of Agarose Gels

Agarose gels were prepared for observing PCR products and restriction digests. 1.5 % agarose gels in 1X TBE (0.089 M Tris-borate, 0.089 M boric acid, 0.002 EDTA) (20) were used for PCR while 0.8% agarose gels in TBE were used for restriction digests.

Electrophoresis glass plates were edged with two layers of tape to form the mold. The plate was placed into a cold room (10°C) on a balanced, level platform and allowed to cool. Combs were placed in the mold 1 mm above the glass plate. The agarose and the water were heated until the agarose dissolved and then 5X TBE was added along with ethidium bromide if the gel was to be used for PCR. The mixture was then carefully poured into the mold and allowed to solidify. When solid, a small amount of 1X TBE was poured around the comb and the comb was carefully pulled out. The tape was taken off of the plate and the gel was ready for use. Gels were stored in plastic wrap at 10°C.

Gel Electrophoresis

The gel was placed into an electrophoresis box which was connected to a power supply. One-fifth volume of gel-loading buffer (0.25% bromophenol blue, 40% w/v sucrose) was added to each reaction. Twenty microliters of each reaction were then pipetted into a well of the gel. Gels containing PCR products were run at 180 V for 30 minutes to an hour. Gels containing restriction enzyme digestions were run at 30 V overnight (20).

Staining of DNA

DNA in agarose gels was detected with ethidium bromide. One of two methods was used to expose the DNA to ethidium bromide: 7.5 µl of 10 mg/ml was added to the liquid agarose before gel was poured, or the gel was soaked in a dilute ethidium bromide solution (~200 ng/ml) for 30 minutes and then washed in dH_2O or $CuSO_4$ for 20 minutes

(20). Gels were then placed onto a Fotodyne ultra-violet light box and photographed with a Fotodyne camera with Polaroid 667 film. A 1 second or ½ second exposure was used with an aperature opening of 4.5.

Results

Synthesis of Primers

Table 2: Results of primers synthesis: primers and their concentrations.

Forward Primer	**OD_{260}**	**Concentration**	**Reverse Primer**	**OD_{260}**	**Concentration**
D12S2046F	0.17226	1421 µg/ml	D12S2046R	0.17036	1405 µg/ml
D12S1914F	0.045025	371 µg/ml	D12S1914R	0.18335	1512 µg/ml
D12S950F	0.029515	243 µg/ml	D12S950R	0.18456	1522 µg/ml
D12S1987F	0.01282	105 µg/ml	D12S1987R	0.16562	1366 µg/ml
D12S928F	0.1950	1608 µg/ml	D12S928R	0.04651	383 µg/ml
D12S1867F	0.0902	744 µg/ml	D12S1867R	0.1941	1601 µg/ml
D12S352F	0.1515	1250 µg/ml	D12S352R	0.03172	261 µg/ml
D12S994F	0.1848	1524 µg/ml	D12S994R	0.0182	150 µg/ml
c1C6F	0.1580	1303 µg/ml	c1C6R	0.2426	2001 µg/ml
D12S341F	0.0641	528 µg/ml	D12S341R	0.2923	2411 µg/ml
D12S1455F	0.1174	968 µg/ml	D12S1455R	0.1317	1086 µg/ml
c12C12F	0.14621	1206 µg/ml	c12C12R	0.27695	2284 µg/ml
D12S94F	0.1977	1631 µg/ml	D12S94R	0.006	50 µg/ml
D12S91F	0.2285	1885 µg/ml	D12S91R	0.1475	1216 µg/ml
D12S1397F	0.21863	1804 µg/ml	D12S1397R	0.00853	70.4 µg/ml
D12S1913F	0.073095	603 µg/ml	D12S1913R	0.16077	1326 µg/ml

Sixteen pairs of primers were synthesized successfully on the Gene Assembler. The success of a synthesis was determined by taking the OD_{260} of the reaction products after purification. Synthesis products which showed an OD_{260} of less than 0.005 (> 40 ng/μl) were discarded and the synthesis was repeated until successful (see table 2).

Growth of E. coli and Isolation of Cosmids

Of the 196 clones obtained from Dr. Montgomery, 196 were successfully grown and 179 cosmids were successfully extracted and purified. Ten cosmids DNAs were randomly selected for determination of DNA concentration (Table 3). The average DNA concentration was 1.993 μg/μl.

Table 3: Cosmid DNA isolated from E. coli and its concentrations.

Cosmid	**OD_{260}**	**Concentration**
107G5	0.0750	618 μg/ml
175B12	0.5505	4541 μg/ml
143A5	0.0854	704 μg/ml
53H12	0.1722	1420 μg/ml
96E10	0.2594	2140 μg/ml
111H1	0.2758	2275 μg/ml
189C3	0.3343	2757 μg/ml
206B8	0.1914	1579 μg/ml
66E4	0.1775	1464 μg/ml
160H3	0.2944	2428 μg/ml

PCR Trials

Primer specific PCR was first run with cosmids which were previously mapped to a known marker as controls. Table 4 shows the PCR conditions and minimum amplified band length for each pair of primers. For convenience forward and reverse primers are referred to by the name of the marker that they define. During experimental trials, positive bands were obtained with cosmid 170G6 and primers D12S950, band A; cosmid 128D6 and primers D12S928, band B; and cosmid 214C9 and primers D12S994, band C. These can be seen in figure 20.

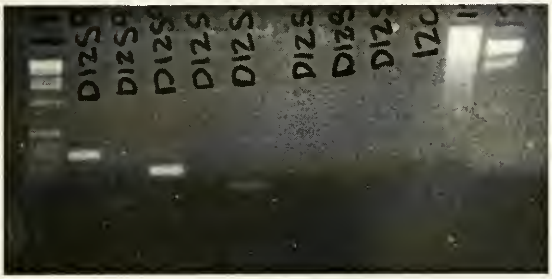


Figure 20: Results of PCR. Lane contents from left to right are as follows: Lane 1, 1 KB ladder; Lane 2, primers D12S950 and cosmid 170G6 (Band A); Lane 3, primers D12S950 and K562 human DNA, Lane 4, primers D12S928 and cosmid 128D6 (Band B); Lane 5, primers D12S928 and K562; Lane 6, primers D12S994 and cosmid 214C9 (Band C); Lane 7, primers D12S994 and K562; Lane 8, primers D12S994 and cosmid c1C6; Lane 9, primers D12S994 and K562; Lane 10, primers c12C12 and cosmid 12C12; Lane 11, λ Hind III digest; Lane 12, 1 KB ladder. bases.

Table 4: STS markers, their optimal PCR conditions, and expected band size. All markers are assumed to be run at 30 cycles unless specified as (35).

STS Marker	Mg++ Conc.	Annealing Temp.	Min. Amp. Length
D12S2046 (932d2-L)	2.0	58° C	0.278 KB
D12S1914 (SHGC-10331)	2.5	62° C	0.091 KB
D12S950 (c170g6)	1.5	58° C	0.280 KB
D12S1987 (SHGC-9936)	1.5	56° C (35)	0.224 KB
D12S928 (c128D6)	1.5	58° C	0.184 KB
D12S1867 (SHGC-5557)	2.5	62° C	0.229 KB
D12S352 (AFM303XD9)	1.5	55° C	0.159 KB
D12S994 (c214C9)	1.5	58° C	0.080 KB
c1C6/T7	2.0	56° C	
D12S341 (AFM294yd4)	1.5	55° C	0.19 KB
D12S1455 (WI-5272)	1.5	56° C (35)	0.177 KB
c12C12/CA	1.5	58° C	
D12S94 (AFM206ze5)	1.5	55° C	0.200 KB
D12S91 (AFM182xf10)	1.5	55° C	0.180 KB
D12S1397 (WI-1025)	1.5	58° C	0.251 KB
D12S1913 (SHGC-10313)	2.5	62° C	0.178 KB

The PCR products were run concurrently with a 1 KB ladder for size determination. The migration distance of each band of the 1 KB ladder was plotted against the log of the band size (figure 21). This plot was used to determine the sizes of the experimental bands. Band A was calculated to have a size of 270 bases which is close to the expected size of 280. Band B has a size of 180 bases which is very close to the

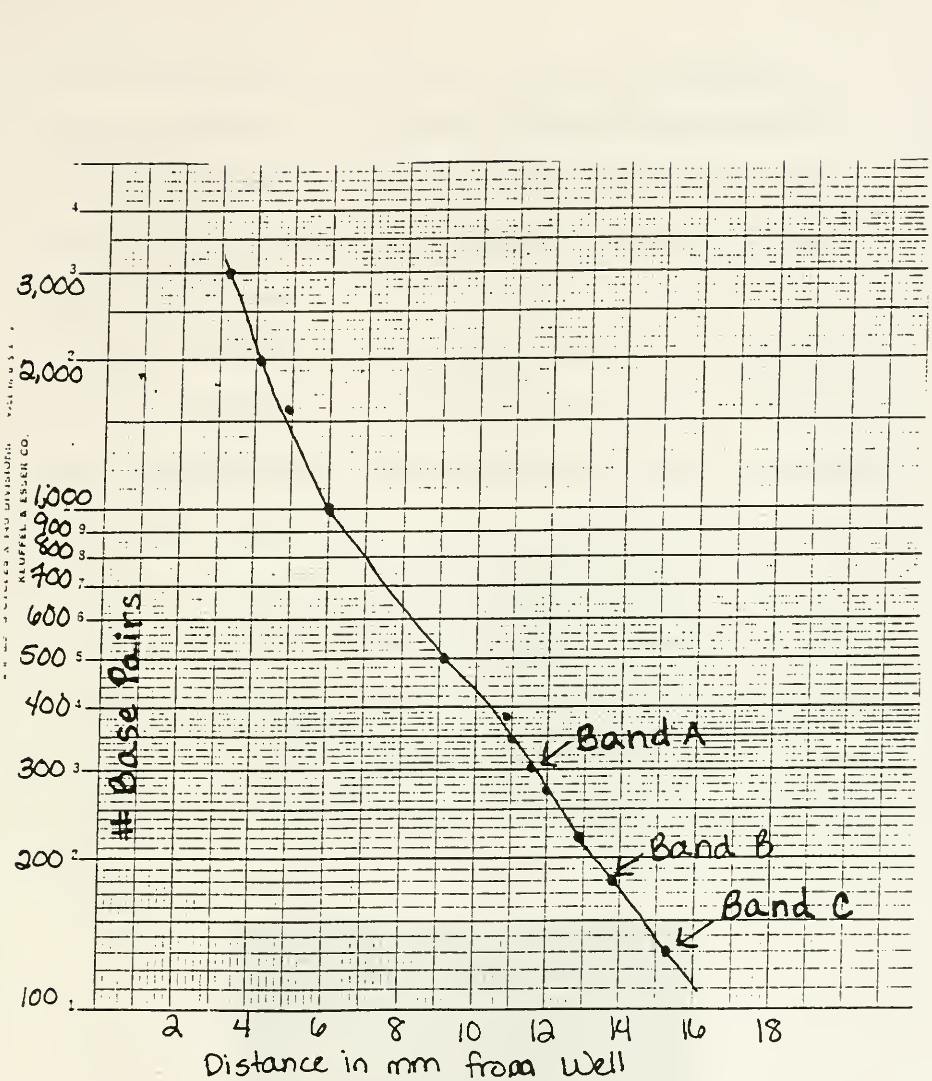


Figure 21: Plot of 1 KB ladder. The size of each band is plotted on the y-axis and the distance traveled in mm is plotted on the x-axis

expected size of 184 bases. Band C has a size of 60 bases which is smaller than the expected size of 80 bases. The size of this band was especially difficult to determine because the small bands of the 1 KB ladder were very faint and difficult to measure. These bands were all expected as each of the cosmids had already been mapped to the marker it was tested against.

Figure 22 shows a band produced by cosmid 128D6 with primers D12S1987. The band has a size of 280 bases and was expected to have a minimum size of 224 bases. This band was unexpected because cosmid 128D6 was already assigned to a marker and it is quite unusual for a cosmid to span two markers. It is unusual because the markers on this map average 100 kb apart from one another. The cosmid 128D6 is under 50,000 bases which means that these two markers must be within 50,000 bases of one another. Marker D12S950 was also tested against 128D6 and found negative. The band showing that cosmid 128D6 was positive for both markers D12S928 and D12S1987 determined not only that the cosmid spanned both markers, but also the exact postion of D12S1987 which was previously known in relation to D12S950 and D12S1914.

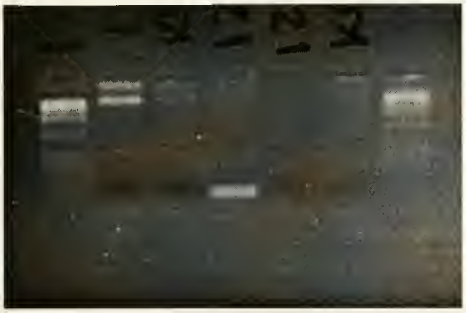


Figure 22: Results of PCR. Lane contents from left to right are: Lane 1, 1 KB ladder; Lane 2, primers D12S1987 and cosmid 170D6; Lane 3, primers D12S1987 and K562; Lane 4, primers D12S1987 and cosmid 128D6; Lane 5, primers D12S1455 and cosmid 128D6; Lane 6, primers D12S1455 and K562, Lane 7, 1 KB ladder.

Pools of cosmids were run against one pair of primers at a time (32). To determine the appropriate number of cosmids to place in a pool, PCR was run with primers D12S950F and D12S950R and three different templates: cosmid 170G6 (known match), pool #1 (5 cosmids including 170G6), and pool #2 (10 cosmids including 170G6). It was found that, although 170G6 by itself gave the strongest band, pool #1 gave a band that was of adequate strength to be detected during experimental reactions (figure 23). Pool #2 gave a faint band which could be seen but which was too light to be of use in experimental conditions as it could easily be overlooked.

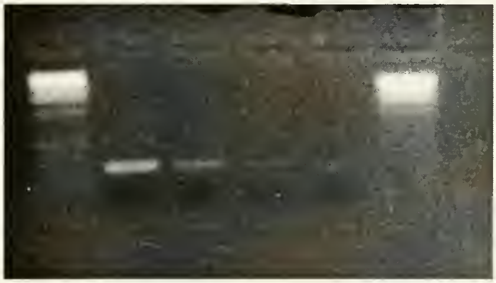


Figure 23: PCR results, lane contents from left to right are as follows: lane 1, 1 KB ladder; lane 2, cosmid 170G6; lane 3, test pool #1 containing cosmid 170G6 and 4 other cosmids; lane 4, test pool #2 contating cosmid 170G6 and 9 other cosmids; lane 5, cosmid 128D6; and lane 6, 1 KB ladder.

Once it was determined that pools of five cosmids were adequate, 58 cosmids were divided into 12 pools and tested against one pair of primers at a time. Table 5 shows pools 1-12 and which cosmids they contain. Pools 1-12 were tested against 7 different primer pairs and 2 reactions gave faint bands. These were primers D12S994 and c1C6/T7. D12S994 showed bands in pools #8 and #12 (figure 24). These cosmids were then run individually against the primers three different times, each time with no positive

bands. Marker c1C6/T7 gave faint bands in pools #5 and #12 (figure 25). The cosmids in those pools were then run against c1C6/T7 two times with no positive bands.

Table 5: Pools 1-12 and the cosmids they contain.

Pool #	Cosmids	Pool #	Cosmids	Pool #	Cosmids
1	56B11	5	123E4	9	175B12
	64E1		126A6		177H10
	66G2		127C3		187A10
	66H7		134A9		189D12
			136G12		197F12
2	67F1	6	142C12	10	198E4
	70D9		143A5		201D6
	79E12		143E9		203B5
	83B2		145D9		209H11
	84E3		147C6		210A7
3	95C12	7	147C7	11	213E4
	97D6		153A9		218A7
	107B9		160E6		219D4
	107G5		160G3		220E1
			160H1		221F2
4	109E1	8	160H3	12	221H6
	112A7		164G4		224G7
	112C8		167F1		226D2
	112H6		168G6		229F11
	119E1		169D10		231A4

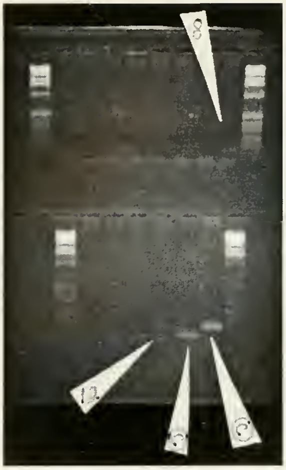


Figure 24: Results of PCR showing faint bands in pools 8 and 12. Bands marked C are control bands. The band on the right is a control with chromosome 12 and the band on the left is a control with cosmid 214C9.

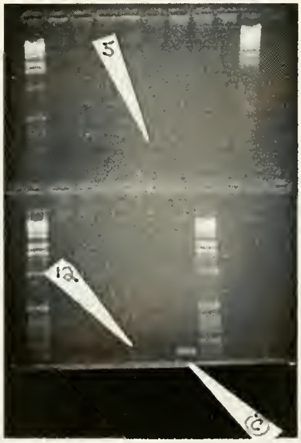


Figure 25: Results of PCR showing faint bands in pools 5 and 12

The effect of the concentration of $MgCl_2$ on the accuracy of the PCR reaction was also determined (figure 26). Three reactions were run with chromosome 12 DNA as template and primers D12S352. 1.5 mM MgCl2 gave a single band of appropriate size,

3.0 mM MgCl2 gave five bands with some smears between bands, and 15mM $MgCl_2$ showed a definite smear with two strong bands, neither of which was the control.

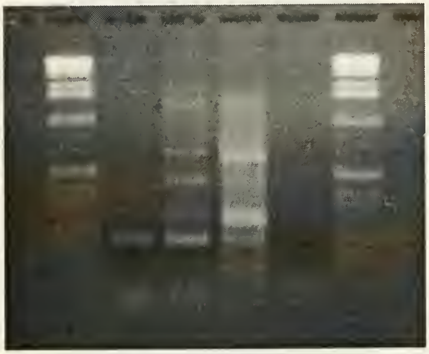


Figure 26: Results of PCR, lane contents are as follows: lane 3, 1 KB ladder; lane 4, 5, and 6, primers D12S352 with chromosome 12 DNA and Mg++ concentrations of 1.5 mM, 3 mM, and 15 mM respectively; lane 7, primers D12S1987 with chromosome 12; lane 8, 1 KB ladder.

Restriction Digestions:

Of the five restriction enzymes originally tried, *Eco*R I and *Bam*H I were found to be most practical because of the number of bands produced, their ability to work well together for double digestions and because they were less expensive than some of the other enzymes. They were therefore used as the primary restriction enzymes. Seven out of the 10 cosmids which were in pools #8 and #12 (possibly positive with D12S994) were digested singly with both *EcoR*I and *BamH*I (figures 27, 28). These cosmids were 167F1, 168G6, 169D10, 221H6, 224G7, 226D2, and 231A4. Table 6 shows the sizes of the cosmids which were digested. This was accomplished by sizing each visible band and taking the sum of all of the bands.

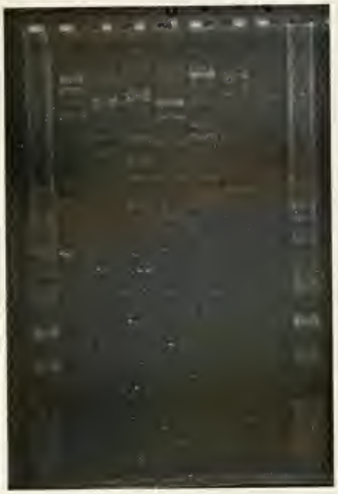


Figure 27: Digestion of cosmids with EcoRI. Lane contents from left to right are as follows: 1 KB ladder, 167F1, 168G6, 169D10, 221H6, 224G7, 226D2, 231A4, 1 KB ladder.

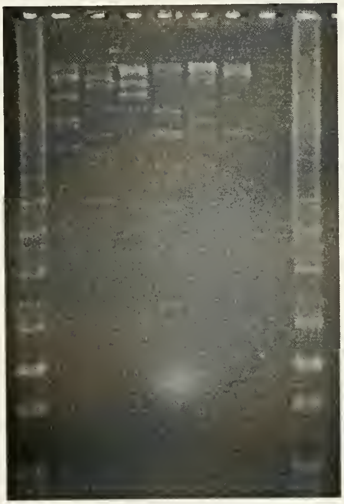


Figure 28: Digestion of cosmids with BamHI. Lane contents are as follows: 1 KB ladder, 167F1, 168G6, 169D10, 221H6, 224G7, 226D2, 231A4, 1 KB ladder.

Table 6: The sizes of cosmids which were digested with EcoRI and BamHI. These were determined by plotting the 1KB ladder bands vs. distance from the well and fitting each of the cosmid bands to the curve. All digested bands were added for a minimum size for the cosmid. Each cosmid may actually be larger due to small digestion fragments which were too small to detect or to multiple bands of the same size.

Cosmid	Minimum Size
167F1	48,100 bases
168G6	42,750 bases
169D10	43,900 bases
221H6	42,550 bases
224G7	63,350 bases
226D2	46,500 bases
231A4	did not digest

Tables 7 and 8 show the results of common bands in the EcoRI and BamHI digestion respectively. A double digestion was also performed and the results of common bands are shown in table 9.

Table 7: Results of digestion with EcoRI. Starred boxes represent positive bands with the size in bases at the top.

	5000	8800	1200	1500	3500	5500
167F1			**********	**********		
168G6	**********	**********	**********			
169D10			**********			**********
221H6	**********	**********	**********	**********		
226D2			**********	**********	**********	
224G7					**********	**********

Table 8: Results of digestion with BamHI

	10000	3700	5000	3900
167F1	**************			
168G6		**************	**************	
169D10	**************			
221H6			**************	**************
226D2		**************		
224G7				**************

Table 9: Results of the double digestion. Sizes of bands in common from the single digestions are shown at the top along with the enzyme which cut them. Yes means that the band was still present after the double digestion, while no means that it was no longer present. ? represents that the data was difficult to interpret.

	Eco 5000	**Eco 1900**	**Eco 8800**	**Eco 1200**	**Eco 1500**	**Eco 3500**	**Bam 5000**	**Bam 3700**	**Bam 3900**
168G6	yes	no	yes	no	yes?		Yes	no?	?
221H6	yes	no	yes	no	yes?		Yes	no?	?
224G7				no	?				

The results of the single *EcoRI* digestion were analyzed and contig 1 was put together. The results of the BamHI digestion were then applied to the contig according to their best fit. Contig 1A was then constructed in order to remove some of the inconsistencies found in the original contig.

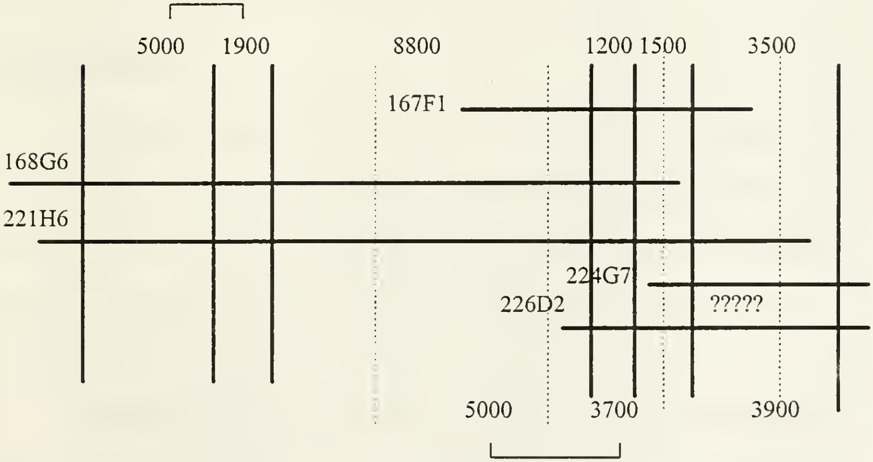


Figure 29: Contig #1. The solid horizontal lines represent the cosmids of the contig. The solid vertical lines represent EcoRI restriction sites within the cosmids and the dotted vertical lines represent BamHI restriction sites. Sizes of the EcoRI bands are shown along the top with BamHI sizes along the bottom. Cosmid numbers are next to each cosmid. ??? represent the data that cosmid 226D2 does not have a 3900 base BamHI band.

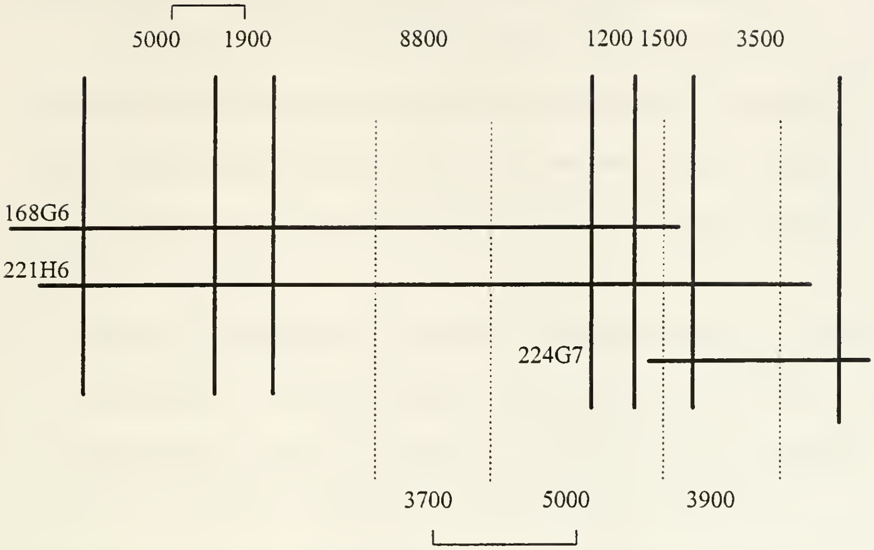


Figure 30: Contig #1A. Modified from contig 1.

Hybridization:

To determine whether the hybridization kit's components were working, a control hybridization procedure was performed using a serial dilution of Schleicher and Schuell control target, pUC19 DNA. Schleicher and Schuell's control labeled oligonucleotide probe and colorimetric detection system were also used. This hybridization was performed with no positive results except for some non-specific binding.

Analysis began with testing of the hybridization procedure and the components of the kit which played an important role in each step of the procedure. Possible sources of error were incorrect annealing temperature, failure of the DNA to attach to the membrane,

failure of the detection substrate, failure of detectable enzymatic activity, incorrect labeling of the probe, and failure of the probe to hydrogen bond with the DNA.

The first step in the hybridization is the determination of the T_m or melting temperature. The annealing temperature is then determined to be 20°C cooler than the T_m. The annealing temperature for the control reagents provided with the kit is given as 42°C. It was therefore known that the annealing temperature was correct and could not be the cause of the problems.

The next step of the hybridization procedure to be tested was the attachment of the DNA to the membrane. Ultraviolet radiation had been used on all previous attempts to attach the DNA to the membrane. The other procedure which can be used is to bake the membrane in an oven at 80°C for one hour after the DNA has been placed on the membrane. This was performed and hybridization was run jointly on two membranes, one which had been exposed to UV radiation and one which had been baked in an oven. No results were obtain in either case. The oven baked membrane did however show a great deal more non-specific binding (figure 31).

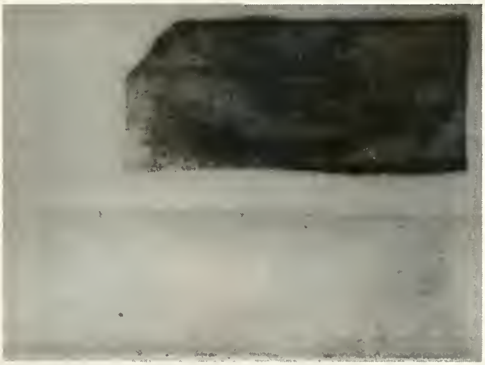


Figure 31: (a) Results seen of a control hybridization with a membrane which was prepared with UV radiation. (b) Increased non-specific binding on a membrane which was prepared by baking at 80°

The detection system was the next component of the kit to be analyzed. All previous attempts at hybridization were performed with the colorimetric detection system. The chemiluminescent system is much more sensitive and was therefore attempted, but no results were obtained. A new control procedure, using the chemiluminescent detection system was tried to see if the detection substrate and enzyme were of good quality. Schleicher and Schuell λ phage DNA which was pre-labeled with psoralen-biotin was obtained, blotted in a serial dilution onto a membrane, and attached through the use of UV radiation.

The membrane was then blocked, incubated with streptavidin-alkaline phosphatase, and then processed in the chemiluminescent detection step. Positive results were obtained (figure 32) for the following amounts of DNA after 3 hours of incubation against the film: 2.5 ng, 500 pg, and 100 pg. The film was replaced and allowed to incubate overnight. Additional positive results for 20 pg were obtained with the lengthened exposure time (figure 32). This procedure was repeated with the colorimetric detection system. Positive results were obtained for only the 2.5 ng concentration of DNA (figure 33).

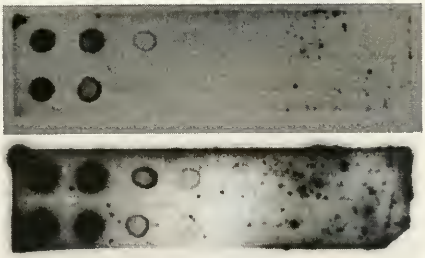


Figure 32: Results of Schleicher and Schuell labeled λ phage DNA control after incubation with film for 3 hours (top) and overnight (above).

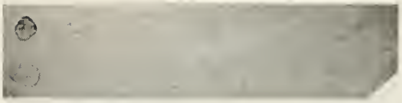


Figure 33: Results of the Schleicher and Schuell labeled λ phage DNA control with the colorimetric detection system.

The next step of the hybridization procedure to be tested was the labeling of the DNA to be detected. Unlabeled λ DNA was obtained from Schleicher and Schuell. The labeling procedure was attempted on the DNA and then processed through the chemiluminescent detection system. Positive results were obtained after 3 hours incubation with film for DNA concentrations of 2.5 ng, 500 pg, 100pg, and 20 pg. Additional positive results were obtained after overnight incubation for a DNA concentration of 4 pg (figure 34).

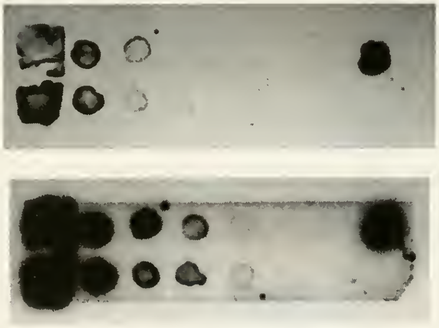


Figure 34: Results of the self-labeled λ phage DNA control hybridization. Membranes were exposed to film for 3 hour (top) and overnight (above) exposures.

The only remaining step of the hybridization procedure to be tested was the actual hydrogen bonding between the probe and the target DNA. A control was tried which was known to be optimal for hydrogen bonding. Cosmid 170G6, which was known to be positive for marker D12S950, was blotted on to a membrane in a serial dilution from 9 µg

to 232 ng. Hybridization was then attempted with primer D12S950F. No positive results were obtained.

The next trouble shooting step was to physically submerge the hybridization bag in the water bath to ensure that the temperature could not deviate from the correct annealing temperature. This was attempted with both the cosmid 170G6 with primer D12S950F and with the control dot blot with control primers. No positive results were obtained from either procedure.

One membrane which had been processed through the hybridization procedure was soaked in dilute ethidium bromide for fifteen minutes and then exposed to UV radiation in an attempt to detect double stranded DNA on the membrane.. No results were obtained.

One more attempt was made to ensure that proper hydrogen bonding was occurring between probe and target. This was the use of a larger probe to make hydrogen bonding between probe and target more stable. PCR was performed on cosmid 170G6 and primers D12S950 and the PCR products were blotted onto a membrane from 1 μg to 3 ng. Hybridization was performed on the membrane with primer D12S950F. No positive results were obtained with this procedure.

Discussion

Synthesis of Primers

Successful synthesis of primers was accomplished on the Gene Assembler®. Although some of the primers had low yields according to the gene assembler's read out, adequate quantities of completed oligo were detected with the spectrophotometer as being present after the first synthesis. The few primer syntheses which did not have adequate concentrations were repeated and were successful on the second try. The main difficulty with syntheses was the aging of the amidites. Amidite dGTP was especially difficult to keep from degrading and usually was the cause of any low yields obtained.

Isolation of cosmids from E. coli

Cosmid DNA was successfully isolated from 179 of the 196 cosmids. Problems that were encountered in the 17 which were not isolated were problems in the growth of the cosmids. Cosmids were originally grown on plates sent by Dr. Kate Montgomery and were subsequently re-plated to form single colonies on Kanamycin plates. Cosmids for which growth on new plates was not obtained had been stored at 10° C for several months and may have dried out. It is also possible that some of these bacteria could have deleted their cosmids and therefore lost their resistance to the Kanamycin in the agar.

Cosmid DNA which was isolated later in the project tended to give larger pellets when precipitated than cosmid DNA which was isolated early in the project. These cosmids also had higher concentrations of DNA when measured on the

spectrophotometer. Although this could have been caused by increased efficiency in the procedure due to practice, it is likely that it came instead from the addition of the Phase-Lock Gel™ into the protocol. The Phase-Lock Gel™ made extraction from the phenol-chlorophorm-isoamyl-alcohol easier and more accurate by forming a solid interphase between the DNA solution and the phenol. A dramatic increase in the size of the pellets was observed after using this product.

PCR Trials

The polymerase chain reaction was successful in the testing of cosmids which had been previously mapped. The success of controls (previously mapped cosmids) determined that PCR conditions and the reagents used were all in working order. The assignment of marker D12S1987 to a place on the chromosome within cosmid 128D6 will be of great use to researchers working on this section of the chromosome. It will help in the ordering of recombinant DNA not only by showing its own location but also by narrowing down the placement of markers D12S1914 and D12D950. The assignment of cosmid 128D6 to two different markers also is exciting because it means that these markers must be close to one another. It is unusual for a cosmid to span two markers because the average distance between markers in this area is 100,000 bases. Having two markers close to one another will make formation of a contig in that area much easier.

PCR reactions which used both control template DNA (K562, chromosome 12, and YAC 932d2) and unassigned cosmids gave results which were not reproducible. The chromosome 12 DNA seemed to be the least reproducible out of the three different

control templates. PCR reactions run with cosmids which had not been assigned to an exact location were unsuccessful, producing only 2 out of 12 reactions which produced any experimental bands and these were not reproducible.

The 179 cosmids are known to be located within an area containing sixteen markers and therefore it would be expected that there would be approximately 11 cosmids per marker. Since the first PCR reactions were run against 12 pools of 5 cosmids each it would be reasonable to expect to see one to four bands in each of the pooled reactions and 3 or 4 bands when run against individual cosmids (4). The two reactions which did produce bands against pools both showed two bands. This number is consistent with expectations.

However, after running PCR against individual cosmids, one would expect to see at least 2 definite strong bands and instead none were visible (25). One possible reason that PCR may not have worked would be that the cosmid DNA was not "clean" enough and that some contaminant could have prevented amplification. This seems very unlikely however because the cosmid DNA is digested well by restriction enzymes which also require very clean DNA (20). Additionally, the PCR reaction which produced two faint bands against primers D12S994 also gave different sized bands in the two control lanes which suggests that the PCR was not correct.

The possibity of an error due to incorrect PCR conditions was ruled out by the magnesium concentration experiment. While it is true that the wrong concentration of magnesium could give a band shift, it would also give many other extraneous bands. Since only one band of the wrong size was present it is unlikely that it is due to inappropriate condition. A more likely cause is that the band shift was due to a restriction

fragment length polymorphism (RFLP). RFLPs are allelic differences located between the two primers so that one allele has a larger area amplified.

Restriction Enzyme Digestions:

Restriction enzyme digestions were very successful. Conditions were optimized by increasing the amount of DNA to be digested so that bands were brighter and more distinct due directly to an increased quantity of DNA within the gel. This increase in DNA concentration especially helped in the observation of small bands which tend to be difficult to see due to the smaller quantity of DNA present and their tendency to diffuse through the gel more readily. Both single and double digestions were well defined showing bands which were easily analyzed and compared.

Construction of a contig:

From single *EcoR*I digestion results contig 1 was put together and is shown in figure 31. The *BamH*I results were then analyzed and fit to this contig. The *Bam*HI results created many inconsistencies in the contig and it was determined to be at least partially inaccurate. One of the inconsistencies is that 226D2 does not have the 3900 base *BamH*I band and yet it does have the 3500 *EcoR*I band. This caused 226D2 to be deleted from contig 1A (figure 32). Also 167F1 contains the 1200 and 1500 base bands of *EcoR*I but does not have the 3700 base *BamH*I band. This could be compensated though by the reversal of the 5000 and 3700 base BamHI bands. However, 167F1 is probably still inaccurate since 167F1 is at least 35,800 bases in length (this was

determined by plotting a 1 KB ladder which was run in an adjacent well and then determining the size of each of the bands from 167F1. (The sizes totaled 35,800). In order for 167F1 to fit on the contig it would have to be less than 15,000 bases long. Cosmid 167F1 was also deleted from contig 1A.

Out of the remaining cosmids, 168G6 and 221H6 appear to have a large overlap and 224G7 appears to have a small overlap with 221H6. One of these cosmids must be incorrectly placed on this contig. This is because the cosmid 168G6 has a minimum size of 42,750 bases and cosmid 221H6 has a minimum size of 42, 550. If these two overlapped as shown in contig 1A then one would expect that they would also overlap to the right of the contig since the area shown accounts only for approximately 17000 bases. This however is not the case because the bands shown are the only bands which they shared in common. It is possible that one cosmid continues to the right while the other continues to the left. If this is true, then cosmid 224G7 no longer fits on the contig since it shares only one band with 221H6.

The extensive overlap of 168G6 and 221H6 suggests that these two form the beginning of a contig. Certainly more data is needed to make the contig more complete and accurate.

Hybridization:

Hybridization of oligonucleotide probes to cosmid DNA has been, up to this point, unsuccessful. There are 6 main components of the hybridization procedure and kit which can be and were tested. These are the determination of T_m and optimal annealing temperature, the attachment of DNA to the membrane, the detection system, the enzyme

(streptavidin-alkaline phosphatase), the labeling system, and the hydrogen bonding between probe and target DNA.

The controls supplied with the oligo hybridization kit have an optimal annealing temperature of 42°C. This information was supplied with the kit and there can be no discrepancy as to the validity of this information. It is therefore possible to rule out the determination of T_m and annealing temperature as the cause of the problem with the hybridization procedure.

The next component of hybridization is the attachment of DNA to the membrane. A comparison of baked membrane to UV radiated membrane did not show if either procedure worked since negative results were received with both membranes. However, it was seen that by baking the membrane a larger amount of non-specific hybridization, or background noise was produced. This implies that out of the two methods, the UV radiation method is preferable assuming both methods are comparable in their ability to crosslink the DNA to the membrane. The pre-labeled λ DNA control demonstrated that UV radiation does indeed crosslink DNA to a membrane. The positive results obtained could not have been possible if the DNA had not been successfully and permanently attached to the membrane.

The chemiluminescent detection substrate and the streptavidin-alkaline phosphatase were also shown to be working in the pre-labeled λ DNA control. The streptavidin-alkaline phosphatase had to be active in order for the substrate to give off light which exposed the film. Therefore, the enzyme and substrate of the detection system have been shown to be active and producing results. The colorimetric substrates

were also shown to work although with a lower sensitivity. While the chemiluminescent substrate was sensitive all the way to 20 pg, the colorimetric substrate was only sensitive to 2.5 ng. This almost 100 fold difference in sensitivity demonstrated that the colorimetric substrate, while being active, was not of use to this project which requires much greater sensitivity.

The other component of the kit which was proven in working order was the psoralen-biotin labeling system. This was demonstrated by the unlabeled λ DNA control. The λ DNA was labeled with the psoralen-biotin oligo labeling procedure and then blotted onto a membrane in a serial dilution as above. The membrane was then processed with the chemiluminescent detection system and positive results were obtained. The results of this control showed greater sensitivity than those in the pre-labeled λ DNA control. While the pre-labeled results showed a sensitivity of 100 pg after a 3 hour exposure and 20 pg after an overnight exposure, the self-labeled results showed a sensitivity of 20 pg after just 3 hours and a 4 pg sensitivity after an overnight exposure. These results seem to agree with results by Pollard-Knight et al.(27).

The final component to be tested was the hydrogen bonding between the probe and the target. Four control experiments were performed in an attempt to show that hydrogen bonding was occurring. This first was a hybridization between cosmid 170G6 and primer D12S950F. This was a primer which was known to amplify a part of the cosmid in PCR and therefore must have a complementary sequence. If the conditions were correct for hybridization the duplex would have formed and formed a detectable signal. The lack of results from this only shows that either hydrogen bonding did not

occur under the conditions used or that not enough hydrogen bonding occurred to form a signal.

Physical submergence of the hybridization bag was tried next. The hybridization bag should have been submerged on all trials but was not tried until late in the trouble shooting procedure. The submergence of the bag ensures that the temperature inside the bag stays extremely stable and accurately at the annealing temperature. A glass bottle sat on top of the bag within the water bath during the entire hybridization and pre-hybridization periods. Hybridization was attempted both with the cosmid 170G6/primer D12S950F and with the kit controls. Lack of results still supported that hydrogen bonding was not occurring.

A membrane which had been processed through the entire hybridization process was then soaked in ethidium bromide. The purpose was to try and visualize any double stranded DNA either from a template/probe duplex or from cosmid DNA which had not been completely denatured. No results were obtained when the membrane was exposed to UV radiation. This experiment was not likely to be positive even if duplex DNA did exist. The reason behind this is that if ethidium bromide could detect the probe/target duplex then the hybridization kit would not have had to contain such an elaborate detection system. This experiment did however add support to the possibility that hydrogen bonding was not occurring.

The final experiment tried was an attempt to strengthen the signal by increasing the relative amount of hybridization region by reducing the non-hybridization region of the template DNA. PCR was performed with cosmid 170G6 and primers D12S950. The resulting products were used as the target and the primer D12S905F was used as the

probe. The lack of results from this procedure strongly supported the idea that hydrogen bonding was not occurring.

By knowing that all of the components of the kit are working except for the actual hybridization step, it can be determined that the failure to achieve results was due to failure to create adequate hydrogen bonding for the detection of the signal. The sensitivity achieved with the self-labeled λ DNA control suggest that very little or no hydrogen bonding must have occurred. One of the causes of failure to hydrogen bond could be contributed to inadequate denaturation of the template. This is highly unlikely since an established protocol was followed.

Another possibility is that the equipment used is inadequate. This may be possible because Sweet Briar does not have all of the equipment that laboratories usually use for hybridizations. It was suggested by one of the technical service workers that making dot blots by hand may cause the DNA to become too diluted on the membrane due to its “wicking” properties and that instead a dot blotting manifold should be used. This seems unlikely since a handmade dot blot was adequate in the λ phage controls. The controls did however show the typical “halo” appearance which is caused by the liquid being carried to the outside of the drop as it dries so this cannot be completely ruled out.

Secondly many people now use a hybridization oven instead of a water bath. By performing hybridizations in a hybridization oven, the sample receives better agitation which would help the probe “locate” the target sequence. The experiments above relied mostly on diffusion to allow the probe to come in contact with the target DNA. To counteract this lack of agitation an experimental hybridization should be tried with the

probe being exposed to the target for a longer period of time such as overnight. This would allow for the probe to completely diffuse through the sample and increase the chances of the probe locating the target.

Another experiment which should be tried is the hybridization of the PCR product (from cosmid 170G6 and primers D12S950) to the cosmid 170G6. By having a larger probe, the hybridization duplex would be much more stable and would therefore be more likely to anneal tightly to the correct sequence. The next step in this project should be to try this hybridization of the PCR product to the cosmid with an overnight hybridization step. If the result is negative then two possibilities will have been ruled out. However if the result is positive then the two factors can be tested independently to determine which one (or possibly both) made the difference. It is worthwhile to try and get hybridization working because this method allows screening of over 200 cosmids in one trial, saving months of work when compared to PCR.

Overall, the beginning of a contig around Marker D12S994 has been started with the two cosmids 168G6 and 221H6. This contig can be continued with additional screening of cosmids with either PCR or hybridization. It may also be possible to extend this contig by examining digests of cosmids which are positive for marker c1C6 which also had positive bands when tested against pools of cosmids. This marker happens to be next to marker D12D994 on the chromosome. Therefore these two markers may very well have some overlapping cosmids. Once a contig of this area is formed, the restriction digest fragments can be isolated and will be sent to AECOM for sequencing.

References

1. Beltz, G., Jacobs, K., Eickbush, T., Cherbas, P. and Kafatos, F. (1983) "Isolation of Multigene Families and Determination of Homologies by Filter Hybridization Methods" *Methods in Enzomology* **100**:266-285.

2. Berg, P., and Singer, M. (1992) Dealing with Genes University Science Books: California.

3. Bonner, T., Brenner, D., Neufeld, B., and Britten, R. (1973) "Reduction in the Rate of DNA Reassociation by Sequence Divergence" *Journal of Molecular Biology* **81**:123-135.

4. Chumakov, I., LeGall, I., Billault, A., Ougen, P., Soularus, P., Guillou, S., Rigault, P., Hui, H., DeTand, M., Barillot, E., Abderrahim, H., Cherif, D., Berger, R., Paslier, D. and Cohen, D. (1992) "Isolation of chromosome 21-specific yeast artifictial chromosomes from total human genome library" *Nature Genomics* **1**: 222-225.

5. Cohen, D., Chumakov, I., and Weissenbach, J. (1993) "A first generation map of the human genome" *Nature* **366**: 698-701.

6. Collins, F. and Galas, D. (1993) "A New Five-Year Plan for the U.S. Human Genome Project" *Science* **262**: 43-46.

7. Cooper, Necia Grant. (1994) The Human Genome Project: Deciphering the Blueprint of Heredity. University Science Books: California.

8. Couch, F., Castilla L., Xu, J., Abe.,K., Welcsh, P., King, S., Wong, L., Ho, P., Merajver, S., Brody, L., Yin, G., Hayes, S., Gieser, L., Flejter, W., Glever, T., Friedman, L., Lynch, E., Meza, J., King, M., Law, D., Deaven, L., Bowcock, A., Collins, F., Weber, B., and Chandrasekharappa, S. (1995) "A YAC-,P1, and Cosmid-Based Physical Map of the BRCA1 Region on Chromosome 17q21. (1995) *Genomics* **25**: 264-273.

9. Craig, A., Nizetic, D., Hoheisel, J., Zehetner, G., Lehrach, H. (1990) "Ordering of cosmid clones comvering the Herpes simplex virus type 1 (HSV-1) genome: a test case for fingerprinting by hybridization" *Nucleic Acids Research* **18**:9 2653-2660.

10. Davies, R., (1996) Request for Grant-In-Aid of Research, Sweet Briar College.

11. Fasman, K,. Cuticchia, and Kingsbury, D. (1994) "The GDB™ Human Genome Data Base anno 1994" *Nucleic Acids Research* **22**:17 3462-3469.

12. 5’ → 3’ Inc., 1995/1996 catalog. Boulder Co. 18-20.

13. Houck, C., Rinehart, F., and Schmid, C. (1979) “A Ubiquitous Family of Repeated DNA Sequences in the Human Genome” *Journal of Molecular Biology* **132**: 289-306.

14. Innis, Michael A., Gelfand, and David H., Sninsky, John J., White, Thomas J. (1990) PCR Protocols. Academic Press: San Diego.

15. Isaacs, S., Shen, C., Hearst, J., and Rapopart, H. (1977) “Synthesis and Characterization of new Psoralen Derivative with Superior Photoreactivity with DNA and RNA” *Biochemistry* **16**:6 1058-1064.

16. Krauter, K., Montgomery, K., Yoon, S., LeBlanc-Straceski, J., Renault, B., Marondel, I., Herdman, V., Cupelli, L., Banks, A., Lieman, J., Menninger, J., Bray-Ward, P., Nadkarni, P., Weissenbach, J., Le Paslier, D., Rigault, P., Chumakov, I., Cohen, D., Miller, P., Ward, D., Kucherlapati, R. (1995) “A second generation YAC contig map of human chromosome 12” *Nature* **377**: 321-333.

17. Lee, Thomas F (1991) The Human Genome Project: Cracking the Genetic Code of Life. Plunum Press: New York.

18. Levenson, C., Watson, R., and Sheldon, E. (1990) “Biotinylated Psoratlen Derivative for Labeling Nucleic Acid Hybridization Probes” *Methods in Enzymology* **184**: 577-583.

19. Lodish, Harvey, Baltimore, David, Berk, Arnol, Zipursky, S. Lawrence, Matsudaira, Paul, and Darnell, James. (1995) Molecular and Cell Biology. Scientific American Books: New York.

20. Maniatis, T., Sambrook, J., and Fritsch, E. (1989) Molecular Cloning: A Laboratory Manuel Cold Spring Harbor Laboratory Press: New York.

21. Monaco, A., Lam, V., Zehetner, G., Lennon, G., Douglas, C., Nizetic, D., Goodfellow, P., and Lehrach, H. (1991) “Mapping irradiation hybrids to cosmid and yeast artificial chromosome libraries by direct hybridization of *Alu*-PCR products.” *Nucleic Acids Research* **19**:12 3315-3318.

22. Mullis, K., and Faloona, F. (1987) “Specific Synthesis of DNA *in vitro* via a Polymerase-Catalyzed Chain Reaction.” *Methods in Enzomology* **155**: 335-351.

23. Murray, J., Buetow, K., Wever, J., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V., Sunden, S., Duyle, G., Weissenbach, F., Gyapay, G., Dib, C., Morrissette, J., Lathrop, G., Vignal, A., White, R., Matsunami, N., Gerken, S., Melis, R., Albertsen, H., Plaetke, R., Odelberg, S.,

Ward, D., Dausset, J., Cohen, D., Cann, H. (1994) "A Comprehensive Human Linkage Map with Centimorgan Density" *Science* **265**: 2049-2054.

24. National Research Council. (1988) Mapping and Sequencing the Human Genome. National Academy Press: Washington D.C.

25. Olsen, M., Hood, L., Cantor, D., Botstein, D. (1989) "A Common Language for Physical Mapping of the Human Genome" Science 245: 1434-1435.

26. Pharmacia Biotech's Gene Assembler Special/4 Primers User's Manual

27. Plooard-Knight, d., Simmonds, A., Schaap, A., Akhavan, H., and Brady, A. (1990) "Nonradioactive DNA Detection on Southern Blots by Enzymatically Triggered Chemiluminescence" *Analytical Biochemistry* **185**: 353-358.

28. Saffran, W., Welsh, J., Knobler, R., Gasparro, F., Cantor, C., and Edelson, R. (1988) "Preparation and characterization of biotinylated psoralen" *Nucleic Acids Research* **16**:15 7221-7231.

29. Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K., Erlich, Hl (1988) "Primer-Directed Enzymatic Amplification of DNA with a Thremostable DNA Polymerase" *Science* **239**: 487-491.

30. Sanger, Fredick (1981) "Determination of Nucleotide Sequences in DNA" *Science* **214**: 1205-1210.

31. Schleicher and Schuell Rad-Free® System Rad-Free Universal Oligo Labeling and Hybridization Kit Manual.

32. Sidhu, M., Helen, B., and Athwal, R. (1992) "Fingerprinting Human Chromosomes by Polymerase Chain Reaction-Mediated DNA Amplification" *Genomics* **14**: 728-732.

33. Soeda, E., Hou, D., Osoegawa, K., Atsuchi, Y., Yamagata, T., Shimokawa, T., Kishida, H., Soeda, E., Okano, S., Chumakov, I., Cohen, D., Raff, M., Gardiner, K.,Graw, S., Patterson, D., de Jong P., Ashworth, L. K., Slezak, T. and Carrano, A. (1995) "Cosmid Assembly and Anchoring to Human Chromosome 21" *Genomics* **25**: 73-84.

34. Stallings, R., Torney, K., Hildebrand, C., Longmire, J., Deaven, L., Jett, J., Doggett, N., and Moyzis, R. (1990) "Physical mapping of human chromosomes by repetitive sequence fingerprinting" *Proceedings of the National Academy of the Sciences* **87**: 6218-6222.

35. Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Hallorar, N., Green, P., Thierry-Mieg, J., Aui,L., Sear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, R., Hawkins, T., Ainscough,R., and Waterston,R. (1992) "The *C. elegans* genome sequencing project" a beginning." *Nature* **356**: 37-41.

36. Voet, Donald and Voet, Judith G. (1995) Biochemistry. John Wiley and Sons: New York.

37. Wallace, R., Shaffer, J., Murphy, R., Bonner, J., Hirose, T., and Itakura K. (1979) "Hybridization of synthetic oligodeoxyribonucleotides to ϕχ 174 DNA: the effect of single base pair mismatch" *Nucleic Acids Research* **6**:11 3543-3557.

38. Watson, James D., Tooze, and John, Kurtz, David T. (1983) Recombinant DNA: A Short Course. Scientific American Books: New York.

39. Weaver, R., and Hedrick, P. (1992) Genetics Wm. C. Brown Publishers: Indiana.

7199 5343 19
05•31•01 MAB
INFORMATION CONSERVATION, INC